SFB 649 Discussion Paper 2011-076

Nonparametric Nonstationary Regression with Many Covariates

Melanie Schienle*

* Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

Nonparametric Nonstationary Regression with Many Covariates

Melanie Schienle*

October 2011

Abstract

This article studies nonparametric estimation of a regression model for $d \geq 2$ potentially nonstationary regressors. It provides the first nonparametric procedure for a wide and important range of practical problems, for which there has been no applicable nonparametric estimation technique before. Additive regression allows to circumvent the usual nonparametric curse of dimensionality and the additionally present, nonstationary curse of dimensionality while still pertaining high modeling flexibility. Estimation of an additive conditional mean function can be conducted under weak conditions: It is sufficient that the response Y and all univariate X^j and pairs of bivariate marginal components X^{jk} of the vector of all covariates X are (potentially nonstationary) β-null Harris recurrent processes. The full dimensional vector of regressors X itself, however, is not required to be Harris recurrent. This is particularly important since e.g. random walks are Harris recurrent only up to dimension two.

Under different types of independence assumptions, asymptotic distributions are derived for the general case of a (potentially nonstationary) β–null Harris recurrent noise term ε but also for the special case of ε being stationary mixing. The later case deserves special attention since the model might be regarded as an additive type of cointegration model. In contrast to existing more general approaches, the number of cointegrated regressors is not restricted. Finite sample properties are illustrated in a simulation study.

JEL Classification: C14, C22

Keywords: *multivariate nonstationary time series, recurrent Markov processes, nonparametric estimation, additive models*

*School of Business and Economics, Humboldt University Berlin, Spandauer Str. 1, D-10178 Berlin, Germany. E-mail: melanie.schienle@wiwi.hu-berlin.de. I thank Enno Mammen and Oliver Linton as well as Kyusang Yu for helpful discussions, constant support and encouragement. Furthermore I gratefully acknowledge research support of the German Science Foundation through the Collaborative Research Center 649.

1 Introduction

In this paper we present nonparametric approaches for a general regression set–up under very weak conditions on the covariate process. In particular, the introduced framework provides a first way for nonparametric inference with high dimensional stochastically nonstationary regressors. The setup is motivated by and generalizes cointegration approaches in parametric econometric time series analysis with stochastically nonstationary components. Though results should be of general interest in all application areas where there is no theoretically justified parametric functional form for a regression function, and where an appropriate model involves more than two stochastically nonstationary covariates, of which the fit into standard unit root and long-memory categories might even be controversial. Such settings include among many others e.g. economic exchange rate (Taylor and Sarno, 1998) and demand models (Lewbel and Ng, 2005), but also weather, energy and climate studies (Engle, Granger, Rice, and Weiss, 1986; Harbaugh, Levinson, and Wilson, 2002; Grossman and Krueger, 1995).

In the univariate case, recent literature has established consistency and asymptotic distribution results for nonparametric kernel regression with stochastically nonstationary covariates in the class of (β-) null Harris recurrent processes, which contains mixing processes as subclass but also random walk type processes. Technically, the literature separates into two different strains: See Phillips and Park (1998), Wang and Phillips (2009b) and Wang and Phillips (2009a) for local time, linear embedding techniques and Karlsen and Tjøstheim (1998), Karlsen and Tjøstheim (2001) and Karlsen, Myklebust, and Tjøstheim (2007) for a general Markov coupling time approach.
For multivariate stochastically nonstationary covariates, however, these existing results can generally not be used, since as in the random walk case, such regressors fail to be compoundly recurrent already from dimension two or three on. Corresponding to the standard nonparametric curse of dimensionality where feasibility deteriorates with an increasing number of covariates, we call this a nonstationary curse of dimensionality, where feasibility entirely vanishes at very low dimensions depending on the type of underlying nonstationarity.

In this paper, we introduce a general flexible model framework, where the compound covariate vector can be transient and where it is sufficient for nonparametric type inference if pairs of it satisfy a recurrence property. Weakening assumptions from full dimensional recurrence to pairwise recurrence is key for a multivariate nonparametric regression method without restricting dimensions as smaller than three or nonstationarity as close to stationarity. Reaching this generality in the data, however, the estimated regression function must be additive. Since fully nonparametric estimation is not possible in this case, this seems a mild restriction which still allows for sufficient model flexibility. In this sense, additive estimation countervails two curses of dimensionality: the nonstationary one and as usual the standard stationary one.

Denote observations by subscripts and dimension components by superscripts. In the entire paper we use the shorthand notation $X^{jk} = (X^j, X^k)$. Then given a random design of n joint observations of $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$, we estimate an additive conditional mean function $m : \mathbb{R}^d \to \mathbb{R}$ with component

functions $m_j : \mathbb{R} \to \mathbb{R}$ for $j = 1, \ldots, d$ and scalar m_0 by

$$Y_i = m_0 + \sum_{j=1}^{d} m_j(X_i^j) + \epsilon_i \tag{1.1}$$

for all $i \in \{1, \ldots, n\}$ under suitable identification conditions for $m_j,\ j = 1, \ldots, d$. We assume there is no concurvity, i.e. for $m_1, \ldots, m_d$ nontrivial we cannot have $m_1(x^1) + \cdots + m_d(x^d) = 0$ for all $(x^1, \ldots, x^d)$. The response Y and all univariate X^j and pairs of bivariate marginal components X^{jk} of the covariate vector X belong to the class of β–null Harris recurrent processes.

We introduce a new general backfitting type estimation procedure for the additive model (1.1) which builds on estimation of low dimensional objects only, and for which recurrence of components of the covariate vector is thus sufficient. In the special case of stationary covariates it reduces to smooth backfitting (see Mammen, Linton, and Nielsen (1999)). The asymptotic distribution of the estimation method is derived under different type of independence assumptions on the error, where subcases can be regarded as nonparametric type of cointegration relations. While obtained rates and variances are univariate in form, their driving underlying type of nonstationarity is two-dimensional. Furthermore we investigate to which extent the estimation procedure is robust to model misspecification. We compare its performance to tailored methods improving on robustness in this respect at the price of more restrictive assumptions on the covariates.

For deriving the asymptotic properties of the proposed estimation technique, we show non-standard uniform consistency results for kernel estimators with β–null Harris recurrent processes. These might be of interest on their own.

The paper is structured as follows. In the next section necessary concepts and notations of Markov theory will be presented to keep the paper self-contained. Section 3 introduces framework and estimation techniques and the subsequent section provides convergence and asymptotic results. Extensions are briefly discussed. The finite sample behavior is illustrated in a simulation study in Section 5. The last section concludes. All proofs are contained in the appendix.

2 Motivation and Basic Framework

This section introduces necessary notions and the basic framework for nonparametric regression with multivariate nonstationary covariates. Furthermore form and peculiarities of standard kernel estimators in this general setting are explained motivating the definition of additional tailored versions. These are fundamental in the sequel. For technical details on Markov chain properties we refer to the comprehensive monographs Meyn and Tweedie (1993) and Nummelin (1984).

2.1 Notation and Basic Concepts from Markov Theory

Let $\{X_i\}_{i=1}^n$ be a multivariate aperiodic ϕ–irreducible Markov chain with transition probability P on the state space $(E, \mathcal{E})$ with $E \subseteq \mathbb{R}^d$. The irreducibility ensures that the process does not degenerate to

a subspace of the original space E, i.e. it guarantees the existence of a non-trivial measure ϕ such that for any set $A \in \mathcal{E}$ with $\phi(A) > 0$ it is $\sum_n P^n(x, A) > 0$ for any starting point $x \in E$. As for inference only sets of positive ϕ-measure are of interest, denote the class of non–negative measurable functions with ϕ-positive support by $\mathcal{E}^+$. A set $A \subseteq \mathbb{R}$ is in $\mathcal{E}^+$ if $\mathbf{1}_A \in \mathcal{E}^+$. We need the following short-hand notation: For any non–negative measurable function η and any measure λ define the operator kernel $\eta \otimes \lambda$ by $\eta \otimes \lambda(x, A) := \eta(x)\lambda(A)$, for all $(x, A) \in (E, \mathcal{E})$. For some general operator kernel P denote $P\eta(x) := \int_A P(x, \mathrm{d}y)\eta(y)$ is a function, $\lambda P(A) := \int_{\mathbb{R}^n} \lambda(\mathrm{d}x)P(x, A)$ is a measure and $\lambda P\eta(x, A) := \int_A \int_{\mathbb{R}^n} \lambda(\mathrm{d}x)P(x, \mathrm{d}y)\eta(y)$ is a real number.

Definition 2.1 (small sets and functions)**.** A function $\eta \in \mathcal{E}^+$ is small for a process X with transition probability P if there exist a measure λ, a positive constant $b > 0$ and an integer $m \geq 1$ such that

$$P^m \geq b\eta \otimes \lambda\,. \tag{2.1}$$

A set A is small if $\mathbf{1}_A$ is small. If the measure λ satisfies (2.1) for some η, b and m, then λ is a small measure.

For every ϕ–irreducible Markov chain $(X_i)_i$ there exists a triplet (s, ν, m_0) with a bounded function $0 < s(x) \leq 1$ at all x, and probability measure ν satisfying (2.1) with $b = 1$. For ease of notation, we assume throughout the paper that $m_0 = 1$. Since ν is independent of x, the chain regenerates in small sets and for a process with continuous state space small sets cannot be singletons. In practice, however, detecting small sets from data is a challenge since topological size and form depend on the observed but unknown underlying process. However, every small set is compact if, as for random walks and α-stable processes, X additionally satisfies the Feller property, a continuity assumption on the transition probability operator (see Feller (1971); Meyn and Tweedie (1993)).

We furthermore assume that X or components of it are β–null Harris recurrent. As in a univariate setting, this is the key assumption for nonparametric kernel type inference to be possible (see Karlsen and Tjøstheim (2001); Karlsen, Myklebust, and Tjøstheim (2007)). On top of simple Harris recurrence, which requires a process X to return almost surely to any neighborhood $\mathcal{N}_{x,h} = \{y \mid \|y - x\| \leq h\}$ of any $x \in \mathbb{R}^d$ for any h with $\phi(\mathcal{N}_{x,h}) > 0$, β–null Harris recurrence requires more structure of the average number of recurrences over all sample paths without losing processes of practical relevance (see Chen (2000); Darling and Kac (1957)).

Definition 2.2 (β–null Harris recurrence)**.** The chain (X_i) is β–null recurrent if there exists a small non–negative function f, an initial measure λ, a constant $0 < \beta \leq 1$ and a function L_f which is slowly varying at infinity[1] such that

$$\mathbb{E}_\lambda \left[\sum_{i=0}^{n} f(X_i) \right] \sim \frac{1}{\Gamma(1+\beta)} n^\beta L_f(n) \quad \text{for} \quad n \longrightarrow \infty\,, \tag{2.2}$$

where $\mathbb{E}_\lambda$ denotes the conditional expectation given that the initial distribution of X_0 is λ.

[1] A function L is slowly varying at infinity if $\lim_{\lambda\to\infty} \frac{L(\lambda x)}{L(\lambda)} = 1$ for all x

Note that β is global and characterizes the type of nonstationarity of the chain (X_i) in a single parameter with smaller β indicating more nonstationarity. In particular, β is not specific to the choice of the small function f, which can be easily shown by Orey's theorem (see e.g. Karlsen and Tjøstheim (2001), Lemma 3.1). In one dimension, β–null Harris recurrence is a very weak requirement which allows for a wide class of processes including stationary or positive recurrent processes with but also nonstationary processes of long-memory or unit-root type. Instead of a stationary density, such processes are generally characterized by an invariant measure, which is generally only finite on small sets and otherwise σ-finite. Throughout the paper, we assume that any invariant measure has a density π that can thus be estimated.

Examples 1 (β–null Harris recurrent processes)**.** The class of β–null Harris recurrent processes contains

- for $\beta = 1$: all stationary linear, but also nonlinear time series like nonlinear autoregressive models under certain conditions (see e.g. Example 3.1 in Karlsen and Tjøstheim (2001) and the exponential autoregressive process in Cline and Pu (1999));
- for $\beta = 1/2$: the univariate random walk and various nonlinear threshold models containing scalar unit root components (see Meyn and Tweedie (1993), p. 503 ff and (Myklebust, Karlsen, and Tjøstheim, 2010) for a wide range of examples);
- for $\beta < 1$: ARFIMA(d) models with $d \in (0, 0.5)$, univariate α–stable processes for $1 < \alpha \leq 2$ with $\beta = 1 - \frac{1}{\alpha}$ (see Sato (1999)), multivariate α–stable processes if dimension $d \leq \alpha \leq 2$, in particular the bivariate random walk with $\beta = 0$.

Furthermore scalar or bivariate mean-reverting processes, e.g. the Ornstein–Uhlenbeck process $\mathrm{d}X_t = -aX_t\,\mathrm{d}t + \mathrm{d}W_t$ for $a \geq 0$, are β–null Harris recurrent. General conditions on diffusion models satisfying β–null Harris recurrence are discussed in Höpfner and Löcherbach (2000), Examples 3.5. and Bandi and Phillips (2004) exploiting the explicit linear form of the trend and of the Brownian type stochastic part.

For larger dimensions, however, recurrence amounts to an increasingly harder criterion for fully compound nonstationary processes to satisfy, e.g. independent random walks are only β–null Harris recurrent up to dimension two and transient beyond. Generally, with increasing dimensions beyond two, β–null Harris recurrence gradually excludes most nonstationary processes and thus cannot allow for substantially more generality as standard mixing settings in high dimensional multivariate regression. Thus, as growing dimensions cause nonstationary processes to drop out of the standard estimation framework, this can be regarded as a nonstationary curse of dimensionality which does not just deteriorate finite sample performance of estimators but rules out estimation completely.

Therefore we introduce the class of pairwise β–null Harris recurrent processes.

Definition 2.3 (pairwise β–null Harris recurrence)**.** A multivariate process X is pairwise β–null Harris recurrent if all pairs of components of X are β–null Harris recurrent.

Note that this class is significantly richer than the fully β–null Harris recurrent processes and in particular allows for nonstationarity in any dimension. In particular, it includes the practically important case of any d-dimensional vector of random walks independent of their correlation structure. This is also true for all processes with univariate $\min_{1\leq j\leq d}\beta^j \geq 1/2$. Processes with $\beta^j < 1/2$ are more nonstationary in their univariate recurrence behavior than a random walk but can still be pairwise β–null Harris recurrent under restricted dependence structures.

Denote the sequence of consecutive recurrence times of a β–null Harris recurrent process X in $\mathbb{R}^d$ by $(\tau_k)_{k=0}^{T}(n)$ with $0 \leq \tau_k \leq n$, where the maximal number of recurrences for a given sample size is $T(n) = \max_k \{k : \tau_k \leq n\}$. Note that while any such recurrence time is also a regeneration time for any marginal components of X, recurrences of marginal components might be more frequent depending on dimension and their type of nonstationarity. Therefore we need to distinguish in particular $(\tau_l^j)_{l=1}^{T^j(n)}$ the sequence of recurrence times for the univariate marginal process X^j and $(\tau_l^{jk})_{l=1}^{T^{jk}(n)}$ for the bivariate X^{jk} where the inclusion holds $\left\{\tau_l^{jk}\right\}_l \subseteq \left\{\tau_l^{j}\right\}_l$ for any $0 \leq j,k \leq d$ but not the other way around. These recurrence times allow a split chain decomposition of any ϕ–irreducible Harris recurrent process into blocks of identically distributed parts. This is of major importance for deriving asymptotic results in this setting. All definitions in the following also have marginal analogues. For any $g \in L_\pi^1(\mathbb{R}^d, \mathbb{R})$ it is

$$S_n(g) := \sum_{i=0}^{n} g(X_i) = U_0(g) + \sum_{k=1}^{T(n)} U_k(g) + U_{(n)}(g)\ , \tag{2.3}$$

with $\tau_{-1} := 1$ and blocks

$$U_k(g) = \begin{cases} \sum_{i=\tau_{k-1}+1}^{\tau_k} g(X_i) & \text{when} \quad 0 \leq k \leq T(n) \\ \sum_{i=\tau_{T(n)}+1}^{n} g(X_i) & \text{when} \quad k = (n) \end{cases} \tag{2.4}$$

where $\{(U_k, (\tau_k - \tau_{k-1}))\}_{k=1}^{T(n)}$ are iid with common marginal distribution $U = U(g)$ of U_k, mean $\mu = \mu(g) = \mathbb{E}U(g) = \pi_s(g)$, and variance $\sigma = \sigma(g) = \mathbb{V}U(g)$. The stochastic quantity $T(n)$ plays the role of effective sample size as $T(n)\to\infty$ a.s. for $n \to \infty$. As for a β–null Harris recurrent process, $T(n)$ is on average over all paths of order $n^\beta L(n)$, the actual size of $T(n) \leq n$ a.s. and its distribution over sample paths depend on the regularity β of the underlying process and is not observable. As estimation of β means estimating the tail index of a recurrence time process, the small sample performance of any such estimator will be very poor independent from the chosen procedure. Therefore we directly introduce the observable quantity

$$T_C(n) := \sum_{i=0}^{n} \mathbf{1}_C(X_i) \tag{2.5}$$

for $C \in \mathcal{E}^+$. If C is small for X, $T_C(n)$ and $T(n)$ are asymptotically equivalent in the sense $\frac{T_C(n)}{T(n)} \xrightarrow{\text{a.s.}} c$ with $c > 0$ constant (Remark 3.5. in Karlsen and Tjøstheim (2001)).

2.2 Kernel Estimators and Peculiarities for Multivariate Nonstationary Data

We observe the multivariate pairwise β–null Harris recurrent process X on a fixed bounded set $\mathcal{G} = \mathcal{G}_1 \times \ldots \times \mathcal{G}_d$ of the state space E, where $\mathcal{G}_j \subseteq E_j \subseteq \mathbb{R}$ is bounded for all $j = 1, \ldots, d$. While for

stationary data, the measure of the full dimensional data generating process exists and is bounded by definition on its entire support, for nonstationary data, a restricted bounded support is crucial and sufficient for a continuous invariant measure to be finite. This is easy to see in the case of a univariate random walk where the invariant measure is the Lebesgue measure. It holds for general pairwise β–null Harris recurrent processes, as the entire space can be covered by small sets, on which any of the existing invariant measures π_{jk} is finite (see Meyn and Tweedie (1993)), and for any bounded set there exists a collection of small sets of which a finite number is sufficient to cover it. Thus we work on a bounded support not just for convenience simplifying technicalities such as integration steps in an estimation procedure, but because it is systematically important as a minimal condition for integrals with respect to the invariant measure to exist.

In practice of finite samples, the empirical support of the data is naturally bounded and wlog we can assume the data to be in $[0,1]^d$ after a monotone transformation. But in asymptotic derivations restricting the support of a nonstationary process to be bounded has a systematic impact. The fixed truncation of the support imposes a bias on any estimation procedure as the amount of data outside $\mathcal{G}$ depends on the degree of underlying nonstationarity of the process and might therefore vary along different dimensions and directions. Asymptotically the available amount of data points and actual elements of different marginal component processes within $\mathcal{G}$ might generally differ almost surely even if $\mathcal{G}_j = \mathcal{G}_k$ for $k \neq j$ depending on the type of nonstationarity of the marginal processes. We will work with the following index sets

$$I_{jk}(X^s) = \left\{ i \in \{1,\dots,n\} \,|X_i^s \in \mathcal{G}_s,\ i < \tau^s_{T^{jk}(n)+1} \right\} \tag{2.6}$$

for each $j,\ k \in \{1,\dots,d\}$, s any possible nontrivial subset of the powerset $\mathcal{D}$ of $\{1,\dots,d\}$, and with the convention $\tau^s_{T^s(n)+1} := n-1$. The univariate index set I_j can be obtained from (2.6) with $j=k$ and I_f analogously with scaling according to full dimensional $X \in \mathcal{G}$ instead of X^{jk}. Note that on average over all sample paths the last restriction in (2.6) is binding for all nontrivial $s \in \mathcal{D}$ which do not contain the two elements j,k if for the respective types of nonstationarity it is $\beta^s > \beta^{jk}$. Since recurrence properties generally improve for decreasing dimensionality, this is in particular most likely the case for the univariate marginals with $s \in \{j,k\}$. In order to "balance" estimation with objects based on X^{jk} and X^s the last requirement in (2.6) imposes components X^s to artificially have the same number of effective observations as X^{jk} on $\mathcal{G}$ on average by reducing the number of blocks in the split chain (2.4). Note that observed effective sample sizes are path dependent, thus a speed adjustment with deterministic factors would not work. For a fixed X^s it is $|I_f| \leq |I_{jk}| \leq |I_j| \leq n$ where the absolute value indicates the number of elements in the set on average over all sample paths. If clear from the context we will omit X^s as argument of I in the following. As $T(n)$ and τ are not directly observable, operationalize the index set choice according to (2.6) by $T_C(n)$ as in (2.5) and entry times of C. In a stationary setting, such technical complications are not needed since asymptotically speeds for different components to hit $\mathcal{G}$ and recurrence properties are all of the same order n independent of their dimension.

The basic underlying estimation technique will be kernel smoothing with product kernels. Denote

$$K_{x,h}(X_i) = \prod_{j=1}^{d} K_{x^j,h^j}(X_i^j) \tag{2.7}$$

where each univariate factor satisfies for all $x^j \in \mathcal{G}_j$

$$\int_{\mathcal{G}_j} K_{x^j,h_j}(u^j)\mathrm{d}u^j \quad = \quad 1\,. \tag{2.8}$$

Since $\mathcal{G}_j$ is bounded, a standard way to fulfill (2.8) are boundary modified kernels

$$K_{v^j,h_j}(u^j) = \frac{K(1/h_j(u^j - v^j))}{\int_{\mathcal{G}_j} K(1/h_j(w^j - v^j))\mathrm{d}w^j}\,, \tag{2.9}$$

where in (2.9) K is a standard univariate kernel function which is symmetric about 0, bounded with compact support $\mathcal{S}^j = [-c^j, c^j]$ with $0 < c^j < \infty$ and integrating to 1 on $\mathcal{S}^j$. The kernel K may also depend on j, which is suppressed in the following for ease of notation. Note that for v^j and u^j in the interior $\mathring{\mathcal{G}}_{j,2h_j}$ the modified kernels coincide with standard kernels

$$K_{v^j,h_j}(u^j) = 1/h_j K(1/h_j(u^j - v^j))\,. \tag{2.10}$$

It is $\mathring{\mathcal{G}}_{j,2h_j} := \mathcal{G}_j \backslash \partial\mathcal{G}_{j,2h_j}$ where $\partial\mathcal{G}_{j,2h_j} = \big\{x \big| \, \|x - c\| \leq 2h^j c_j$ for any c from the boundary$\big\}$ is the $2h_j$ ring boundary of $\mathcal{G}_j$ for a kernel with support $\mathcal{S}^j$.

Assumption 1.

1. $K(u^j)$ and $K(u^j) \cdot (u^j)^k$ are Lipschitz-continuous for any $u \in \mathbb{R}$ and any power $k < 2p+1$ with Lipschitz constant $\widetilde{L} > 0$, where p indicates the minimal existing number of partial derivatives of m over all directions.

2. $\mathcal{S}_x^{jk} = x^{jk} \oplus \mathcal{S}^{jk} = \big\{x^{jk} + u | u \in \mathcal{S}^{jk}\big\}$ is small for all $x^{jk} \in \mathcal{G}_{jk}$ where $\mathcal{S}^{jk}$ is the support of the bivariate kernel.

In the remaining part of this section, regard j and k as elements of the power set of $\{1, \ldots, d\}$ for obtaining general definitions from the stated univariate and bivariate ones. For standard kernel density and regression estimators in a nonstationary setting, recurrence frequency $T^j(n)$ and index sets I_j must be carefully in line with dimension and direction of the underlying X^j in $\mathcal{G}_j$. Set

$$\widehat{\pi}_j(x^j) \quad = \quad \frac{1}{T^j(n)} \sum_{i \in I_j} K_{x^j,h_j}(X_i^j)\,, \tag{2.11}$$

$$\widehat{m}_j(x^j) \quad = \quad \frac{\sum_{i \in I_j} K_{x^j,h_j}(X_i^j) Y_i}{\sum_{i \in I_j} K_{x^j,h_j}(X_i^j)}\,, \tag{2.12}$$

and operationalize (2.11) with $T_C^j(n)$ as in (2.5) for an appropriate small set. For β–null Harris recurrent X^j, the two estimators provide pointwise consistent estimates of the invariant measure density and a general nonparametric link function m_j respectively. This has been shown in Karlsen, Myklebust, and Tjøstheim (2007) and Karlsen and Tjøstheim (2001) for univariate X^j, but also holds in dimensions $q > 1$ if the compound q-vector is β–null Harris recurrent, with β^q generally smaller than β^j. Rates of

convergence of (2.11) and (2.12) are driven by the recurrence frequency $T^j(n)$ and the occupation time $\widehat{L}_j(x^j) = \sum_{i \in I_j} K_{x^j,h^j}(X_i^j)$ respectively, acting as effective sample sizes. Asymptotically in both cases, they are on average of size $(n^{\beta^j} h)^{-1/2}$ deteriorating with smaller β^j, which we denote as nonstationary curse of dimensionality. For higher dimensional estimators rates $(n^{\beta^q} h^q)^{-1/2}$ also show the standard curse of dimensionality plus a potentially more severe nonstationary curse through β^q. Though in our estimation method later, we also need basic estimators for objects in direction j or jl with effective sample sizes not greater than in the case jk in order to artificially balance speeds of univariate and different bivariate estimators if necessary

$$\widehat{\pi}_j^{(k)}(x^j) = \frac{1}{T^{jk}(n)} \sum_{i \in I_{jk}} K_{x^j,h_{jk}}(X_i^j) \tag{2.13}$$

$$\widehat{m}_j^{(k)}(x^j) = \frac{\sum_{i \in I_{jk}} K_{x^j,h_{jk}}(X_i^j) Y_i}{\sum_{i \in I_{jk}} K_{x^j,h_{jk}}(X_i^j)} \,. \tag{2.14}$$

Analogously to (2.13) set $\widehat{\pi}_{jl}^{(k)}(x^{jl})$, which is $\widehat{\pi}_j^{(k)}(x^j)$ for $l = j$, and $\widehat{\pi}_{jl}(x^{jl})$ for $l = k$. Define $\widehat{L}_j^{(k)}(x^j)$ as the numerator of (2.13)

$$\widehat{L}_j^{(k)}(x^j) = \sum_{i \in I_{jk}} K_{x^j,h_{jk}}(X_i^j) \tag{2.15}$$

For a full dimensional β–null Harris recurrent process analogues to (2.13) and (2.13) with speed of the compound X are denoted by $\widehat{\pi}_j^f$ and $\widehat{m}_j^f$. The nonstationary character for the estimators in (2.13) and (2.14) is determined by the two–dimensional type β^{jk}. Hence rates of convergence of $\widehat{\pi}_j^{(k)}$ and $\widehat{m}_j^{(k)}$ to π_j and m_j respectively are asymptotically on average of order $(n^{\beta^{jk}} h)^{-1/2}$ - which is univariate in form but of bivariate type of nonstationarity. For $\widehat{\pi}_{jl}^{(k)}$ convergence to π_{jl} occurs at most with rate $(n^{\beta^{jk}} h^2)^{-1/2}$ on average. If $\beta^{jl} < \beta^{jk}$ the slower rate governed by β^{jl} prevails. In the other slowed-down cases set for fix j and k

$$\mathcal{L}_{jk} = \left\{ l \neq j | \widehat{L}_j^{(k)} \cdot \left(\widehat{L}_j^{(l)} \right)^{-1} = O_P(1) \right\} \,. \tag{2.16}$$

This set collects all indices of components which are less or equally nonstationary in pair with X^j than component k is, on the realized path of the underlying data. In particular $1 \leq \lambda_{jk} = |\mathcal{L}_{jk}| \leq d-1$, as $\mathcal{L}_{jk}$ contains at least component k itself.

3 Generalized Smooth Backfitting Estimation (GSBE)

Kernel type estimation of a fully general nonparametric regression setting with more than two nonstationary covariates might be inconsistent. In this case, the compound vector of regressors is generally transient implying that additive functionals such as the kernel estimators do not converge in general. But when restricting the functional form of the regression problem as additive, in most of these settings inference is possible for a specific estimation strategy. In this section, we introduce an appropriate framework of general classes of admissible processes for additive estimation using that low-dimensional subcomponents of transient compound vectors are often recurrent, as illustrated in the last section. In

this setting, we develop corresponding estimation methods which thus allow to circumvent the nonstationary curse of dimensionality. As in the stationary subcase, the proposed estimators also countervail the stationary curse of dimensionality. Though here, improved finite sample performance just appears as a positive side effect of additive estimation, given that for most higher dimensional nonstationary data a more general structural relation cannot be estimated at all.

We develop estimation techniques of smooth backfitting type (Mammen, Linton, and Nielsen, 1999), where the iterative estimation steps entirely consist of low-dimensional invariant density estimators and regression smoothers only. For this class of estimators it is sufficient if corresponding low-dimensional components of the covariate vector are β–null Harris recurrent. Note that other kernel based techniques for additive estimation such as marginal integration (Linton and Nielsen, 1995; Tjøstheim and Auestad, 1994) or two–step local partitioned regression (Christopeit and Hoderlein, 2006) would need full–dimensional invariant measure densities in a pre–step, requiring recurrence of the full-dimensional processes and thus suffering from the same nonstationary curse of dimensionality as fully nonparametric regression. And even in the restrictive class of full dimensional recurrent X, their slow preestimation step would lead to only inferior rates.

Assume throughout this section that the regression model has additive form as in (1.1). Furthermore all mentioned densities of invariant measures and all stated integrals exist, i.e. the regression functions m_j are in the respective weighted L^2–spaces.

As a starting point for estimation and illustration of the later estimation method, assume that estimates of the additive component functions from nonstationary β–null Harris recurrent data also minimize the smoothed sum of squares

$$\sum_{i\in I}\int (Y_i - m_0 - \sum_{k=1}^{d} m_k(x^k))^2 K_{x,h}(X_i)\,\mathrm{d}x \tag{3.1}$$

under the constraints for all $j = 1, \ldots, d$

$$\int_{\mathcal{G}_j} m_j(x^j)\widehat{\pi}_j(x^j)\mathrm{d}x^j = 0\ , \tag{3.2}$$

as in the standard stationary smooth backfitting case (SBE). Minimization of (3.1) and standard kernel calculations lead to the following defining system of integral equations. SBE estimators $(\widetilde{m}_0, \ldots, \widetilde{m}_d)$ solve

$$\widetilde{m}_j(x^j) = \widehat{m}_j(x^j) - \widetilde{m}_{0,j} - \sum_{k\neq j}\int_{\mathcal{G}_k} \widetilde{m}_k(x^k)\frac{\widehat{L}_{jk}(x^{jk})}{\widehat{L}_j(x^j)}\mathrm{d}x^k \tag{3.3}$$

$$\text{with}\quad \widetilde{m}_{0,j} = \frac{\int_{\mathcal{G}_j}\widehat{m}_j(x^j)\widehat{\pi}_j(x^j)\mathrm{d}x^j}{\int_{\mathcal{G}_j}\widehat{\pi}_j(x^j)\mathrm{d}x^j} = \frac{1}{n}\sum_{i=1}^{n} Y_i\ , \tag{3.4}$$

where $\widehat{L}_{jk}$ and $\widehat{L}_j$ are occupation time estimates and $\widehat{m}_j$ is a marginal Nadaraya–Watson pilot estimator as defined below (2.11) and as in (2.12) respectively. Identification and the form of the constant terms $\widetilde{m}_{0,j}$ in (3.4) result from the norming conditions (3.2).

In contrast to the stationary case, however, generally on average the recurrence frequency of X^{jk} might be asymptotically of slower order than for X^j. Therefore the quotient $\frac{\widehat{L}_{jk}}{\widehat{L}_j} = \frac{\widehat{\pi}_{jk}}{\widehat{\pi}_j} \frac{T^{jk}(n)}{T^j(n)}$ in the projection part of (3.3) converges to zero almost surely on average over all sample paths. Thus for a valid estimation procedure effective sample sizes and hence speeds of the involved estimators must be artificially synchronized. Since (3.3) only contains one- and two dimensional objects, the fastest common scale to do so is two–dimensional. Such a procedure appears to be applicable if covariate processes have at least bivariate invariant measures for all pairs of components. We therefore introduce the class of pairwise β–null Harris recurrent processes as defined in Definition 2.3 which is general enough to contain many practically relevant high-dimensional nonstationary processes, which are compoundly transient, but which still allows for consistent nonparametric estimation of an additive structural model as shown in the next section. Up to our knowledge, this framework is new to the literature and the obtained estimation method is the first available procedure for nonparametric estimation with multivariate nonstationary regressors.

For balancing terms in (3.3) we must use potentially slower than standard estimators $\widehat{\pi}_j^{(k)}$, $\widehat{\pi}_{jl}^{(k)}$ and $\widehat{m}_j^{(k)}$ of bivariate nonstationary type β^{jk} as defined in (2.13) and (2.14). Also in the backfitting operator for component j, the impact of other directions on any pair of components containing X^j might now differ depending on respective occupation times of component pairs. To ensure consistency, the procedure must reflect this. Define the generalized smooth backfitting estimates (GSBE) $(\widetilde{m}_j)_{j=1}^d$ for the class of pairwise β–null Harris recurrent regressors X as solutions to

$$\widetilde{m}_j(x^j) = \frac{1}{d-1} \left(\sum_{k \neq j} \left(\widehat{m}_j^{(k)}(x^j) - \widetilde{m}_{0,j}^{(k)} \right) - \sum_{k \neq j} \sum_{l \neq j} \int_{\mathcal{G}_l} \widetilde{m}_l(x^l) \frac{\widehat{\pi}_{jl}^{(k)}(x^{jl})}{\widehat{\pi}_j^{(k)}(x^j)} \mathrm{d}x^l \right) , \tag{3.5}$$

with constants

$$\widetilde{m}_{0,j}^{(k)} = \frac{\int_{\mathcal{G}_j} \widehat{m}_j^{(k)}(x^j) \widehat{\pi}_j^{(k)}(x^j) \mathrm{d}x^j}{\int_{\mathcal{G}_j} \widehat{\pi}_j^{(k)}(x^j) \mathrm{d}x^j} = \frac{1}{T^{jk}(n)} \sum_{i \in I_{jk}} Y_i \, . \tag{3.6}$$

They follow from appropriate analogues of the norming constraints (3.2) on solutions of GSBE equations

$$\sum_{k \neq j} \int_{\mathcal{G}_j} m_j(x^j) \pi_j^{(k)}(x^j) \mathrm{d}x^j = 0 \, . \tag{3.7}$$

Note that asymptotically in the projection part of (3.5) only elements $l \in \mathcal{L}_{jk}$ prevail, while all others vanish. If all pairs of components of X have the same type of nonstationarity, the backfitting equations reduce to

$$\widetilde{m}_j(x^j) = \frac{1}{d-1} \sum_{k \neq j} \left(\widehat{m}_j^{(k)}(x^j) - \widetilde{m}_{0,j}^{(k)} \right) - \sum_{k \neq j} \int_{\mathcal{G}_k} \widetilde{m}_k(x^k) \frac{\widehat{\pi}_{jk}(x^{jk})}{\widehat{\pi}_j^{(k)}(x^j)} \mathrm{d}x^k \, ,$$

since $\lambda_{jk} = d-1$ and $\widehat{\pi}_{jl}^{(k)} = \widehat{\pi}_{jl}$ in this case. In particular, (3.5) fully reduces to (3.3) and the norming constraint (3.7) to (3.2) for the special case of identical one- and two-dimensional scales, i.e. an almost stationary added component in the pair. Thus standard smooth backfitting appears as a subcase of generalized smooth backfitting for sufficiently stationary data.

Numerically, we obtain the generalized smooth backfitting estimates as solution to (3.5) via iteration. For each component j start at an arbitrary initial guess $\widetilde{m}_j^{[0]}$, e.g. the marginal Nadaraya–Watson estimator $\widetilde{m}_j^{[0]} = \widehat{m}_j$. Then denote the rth step iterate of the jth component with $\widetilde{m}_j^{[r]}$. Hence iterate according to

$$\begin{aligned}\widetilde{m}_j^{[r]}(x^j) &= \frac{1}{d-1}\sum_{k\neq j}\left(\widehat{m}_j^{(k)}(x^j) - \widetilde{m}_{0,j}^{(k)} - \sum_{l<j}\int_{\mathcal{G}_l}\widetilde{m}_l^{[r]}(x^l)\frac{\widehat{\pi}_{jl}^{(k)}(x^{jl})}{\widehat{\pi}_j^{(k)}(x^j)}\mathrm{d}x^l - \right.\\ &\quad \left. - \sum_{l>j}\int_{\mathcal{G}_l}\widetilde{m}_l^{[r-1]}(x^l)\frac{\widehat{\pi}_{jl}^{(k)}(x^{jl})}{\widehat{\pi}_j^{(k)}(x^j)}\mathrm{d}x^l\right) \qquad (3.8)\end{aligned}$$

until a convergence criterion is fulfilled. Note that $\sum_{k\neq j}\widetilde{m}_{0,j}^{(k)}$ is only different from zero, when the norming condition (3.7) is violated. If we directly set

$$m_0 = \sum_{j=1}^{d}\frac{1}{d-1}\sum_{k\neq j}\frac{1}{T^{jk}(n)}\sum_{i\in I_{jk}}Y_i\,, \qquad (3.9)$$

the centering term $\widetilde{m}_{0,j}^{(k)}$ can be omitted from the algorithm.

Remark 1. Note that the use of boundary modified kernels in (3.5) and (3.8) is crucial in order to obtain unbiased solutions of GSBE also on the boundary of $\mathcal{G}$. Otherwise the density weight in the backfitting operator should be altered from $\frac{\widehat{\pi}_{jl}^{(k)}(x^{jl})}{\widehat{\pi}_j^{(k)}(x^j)}$ to $\frac{\widehat{\pi}_{jl}^{(k)}(x^{jl})}{\widehat{\pi}_j^{(k)}(x^j)} - \frac{\int\widehat{\pi}_{jl}^{(k)}(x^{jl})\,\mathrm{d}x^j}{\int\widehat{\pi}_j^{(k)}(x^j)\,\mathrm{d}x^j}$ and the centering $\widetilde{m}_{0,j}^{(k)}$ can no longer be omitted from the algorithm but must be used in its original integral form (3.6) which no longer reduces to a constant on the boundary. These modifications are the corresponding general analogues to the ones in the standard SBE setting.

3.1 Adaptive GSBE and Projection Properties

For stationary data, the form of the defining equations of the SBE estimator (3.3) has been motivated via a projection argument (3.1) as the corresponding first order conditions for obtaining the best additive fit to the data in a suitably $\widehat{\pi}$–weighted empirical L_2–norm. This implies that even if the underlying true model is not additive, SBE provides reasonable and controllable estimates as additive projections. But for GSBE, in general admissible processes are only pairwise β–null Harris recurrent such that an invariant measure π of the full dimensional compound process might not exist and the projection property in the general sense cannot prevail. The question, however, is, to what extend and in which sense it can be recovered for general data and how it can be improved upon under which conditions. For simplicity, we use an operator representation of the backfitting problem (3.5) as a Fredholm equation of the second kind in the corresponding (empirically) weighted L^2 Hilbert spaces which will also prove valuable for the presentation of the asymptotic results. Componentwise in $j = 1,\ldots,d$ it is

$$\widetilde{m}_j(x^j) = \frac{1}{d-1}\sum_{k\neq j}(1-\widehat{\Phi}_{jk})\widehat{m}_j^{(k)}(x^j) - \sum_{k\neq j}\sum_{l\neq j}[\widehat{A}_{jl}^{(k)}\widetilde{m}_l](x^j) \qquad (3.10)$$

with centering operator $\widehat{\Phi}_{jk}[\widehat{m}_j^{(k)}] = \frac{\int \widehat{m}_j^{(k)}(x^j)\widehat{\pi}_j^{(k)}(x^j)\mathrm{d}x^j}{\int \widehat{\pi}_j^{(k)}(x^j)\mathrm{d}x^j} = \frac{1}{T^{jk}(n)}\sum_{i\in I_{jk}} Y_i$ and projection operators $\widehat{A}_{jl}^{(k)}$ for $l \neq j$ defined as

$$[\widehat{A}_{jl}^{(k)} m_l](x^j) = \frac{1}{d-1}\int_{\mathcal{G}_l} m_l(x^l)\frac{\widehat{\pi}_{jl}^{(k)}(x^{jl})}{\widehat{\pi}_j^{(k)}(x^j)}\mathrm{d}x^l \ . \tag{3.11}$$

Asymptotically it converges to zero for all $l \notin \mathcal{L}_{jk}$, projecting any function $f \in L^2_{\pi_l}$ onto $L^2_{\pi_j}$. Thus the limiting operator is

$$[A_{jl} m_l](x^j) = \frac{1}{d-1}\int_{\mathcal{G}_l} m_l(x^l)\frac{\pi_{jl}(x^{jl})}{\pi_j(x^j)}\mathrm{d}x^l \quad \text{for } l \in \mathcal{L}_{jk} \tag{3.12}$$

and zero otherwise.

Use vector and matrix notation $\widetilde{m} = (\widetilde{m}_1(x^1), \ldots, \widetilde{m}_d(x^d))^T \in \mathbb{R}^d$ and $\widehat{m} = (\underline{\widehat{m}}_1(x^1), \ldots, \underline{\widehat{m}}_d(x^d))^T \in \mathbb{R}^{d\times d}$ where $\underline{\widehat{m}}_j(x^j) = (\widehat{m}_j^{(1)}(x^j), \ldots, \widehat{m}_j^{(d)}(x^j))^T \in \mathbb{R}^d$ to obtain the simplest operator form of (3.5)

$$(I - \widehat{A})\widetilde{m} \quad = \quad \frac{1}{d-1}\mathrm{diag}\big((\mathbf{1} - \widehat{\Phi})\widehat{m}\big) \tag{3.13}$$

with I the identity and operator matrix entries $((a_{jl})) := -\sum_{k\neq j}\widehat{A}_{jl}^{(k)}$ for all $j \neq k$ and $\mathrm{diag}(\widehat{A}) = \mathbf{0}$, $((\phi_{jk})) = \widehat{\Phi}_{jk}$ for all $j \neq k$ and $\mathrm{diag}(\widehat{\Phi}) = \mathbf{0}$, $((1_{jk})) = 1$ for all $j \neq k$ and $\mathrm{diag}(\mathbf{1}) = \mathbf{0}$. By setting $m_0 = \sum_{j=1}^d \frac{1}{d-1}\sum_{k\neq j}\frac{1}{T^{jk}(n)}\sum_{i\in I_{jk}} Y_i$ the centering term can be omitted and the right hand side reduces to $\widehat{m}^{II} = \frac{1}{d-1}\mathrm{diag}\big(\mathbf{1}\widehat{m})\big) \in \mathbb{R}^d$.

The construction of GSBE focused on the weakest assumptions on the covariate process in terms of stationarity assumptions. Under the admissible generality in the data, however, the backfitting operator cannot keep its full projection characteristic as in the standard stationary case. Corresponding to the pairwise scaling of the algorithm, GSBE as defined in (3.5) yields the best approximation to the data via minimizing the distance in the following (semi)norm on the space of additive functions

$$\frac{1}{d-1}\sum_{k\neq j}\sum_{i\in I_{jk}}\int_{\mathcal{G}}(Y_i - m_1(x^1) - \ldots - m_d(x^d))^2 K_{x,h}(X_i)\,\mathrm{d}x \tag{3.14}$$

in m_j subject to the centering constraints (3.7) for each $j = 1, \ldots, d$. Simple calculations show that the GSBE equations (3.5) are the corresponding first order conditions for these optimization problems (for cross-terms see bias calculations in the appendix). As the metric according to (3.14) is designed in terms of pairs of covariates, it yields a best additive approximation to a general true structural model only in a pairwise sense. Thus for more general underlying models, additional components to pairs are treated as design independent. Such a projection behavior is in the stationary case known from marginal integration.

This projection quality can be improved upon, when the underlying data is restricted to be "less nonstationary". Assuming that all sub–tuples of dimension $\gamma > 2$ of the covariate vector are β–null Harris recurrent, allows to scale the backfitting equations (3.3) according to γ dimensional compound component processes. This enhances the projection character of the obtained estimates in comparison to GSBE as they yield the best additive approximation in a general γ–wise sense even if the true model is not additive. For $\gamma = d$ and scaling according to the full dimensional process, this is the additive projection of a fully general nonparametric relationship at the price of the full nonstationary curse of

dimensionality. In general, such γ–adapted smooth backfitting type estimators do not only admit a much smaller class of processes only, but their rates of convergence are governed by types of nonstationarity of the compound γ–tuples of covariates, which are much slower than for GSBE. Hence for feasibility aspects these modifications are less important than GSBE. See Subsection 4.2 for details.

4 Asymptotic Results

In this section, we state the full set of conditions under which we then derive the asymptotic expansion of GSBE if the underlying true model is additive. For all proofs we refer to the Appendix. Countervailing the nonparametric curse of dimensionality, GSBE is scaled according to bivariate types of nonstationarity. We will see that due to the generality in the data inducing such design, GSBE can only reduce the nonstationary curse of dimensionality up to bivariate types of β, whereas the standard curse of dimensionality can be fully cured by imposing the additive structure.

4.1 Assumptions and Asymptotic Results for GSBE

Assumption 2. 1. X is an ϕ-irreducible aperiodic Markov chain.

2. All pairs X^{jk} are β–null Harris recurrent with parameter β^{jk} for all $j,k=1,\dots,d$, $j\neq k$.

3. All invariant densities π_{jk} exist, are bounded, bounded away from zero and have continuous second partial derivatives on $\mathcal{G}_{jk}$, for $j,k=1,\dots,d$, $j\neq k$.

As pointed out in Section 2, finiteness of the pairwise invariant measures is not restrictive for bounded $\mathcal{G}$. Identification and asymptotic expansion of generalized smooth backfitting estimates (3.5) can be obtained by the following assumptions on any bivariate marginal process X^{jk} of X.

Assumption 3. 1. The compound chain (X^{jk},ϵ) is a ϕ-irreducible β–null Harris recurrent Markov chain with transition probability operator $P_{jk\varepsilon}$ and density $\pi_{jk\epsilon}$ of the invariant measure, where $\pi^{\varepsilon}_{jk}(x^{jk})=\int_{\mathcal{G}_0}\pi_{jk,\varepsilon}(x^{jk},\varepsilon)\,\mathrm{d}\varepsilon>0$ for all $x^{jk}\in\mathcal{G}_{jk}$ and $\pi^{\varepsilon}_{jk}(\mathcal{G}^{jk})<\infty$

2. $\mu_{\varepsilon|jk}(x^{jk})=0$ and $\sigma^2_{\varepsilon|jk}(x^{jk})<\infty$ for all $x^{jk}\in\mathcal{G}_{jk}$ where both quantities are defined with respect to invariant measures $\mu_{\varepsilon|jk}(x^{jk})=\int\epsilon\frac{\pi_{jk\epsilon}(x^{jk},\epsilon)}{\pi^{\varepsilon}_{jk}(x^{jk})}\,\mathrm{d}\epsilon$ and $\sigma^2_{\varepsilon|jk}(x^{jk})=\int\epsilon^2\frac{\pi_{jk\epsilon}(x^{jk},\epsilon)}{\pi^{\varepsilon}_{jk}(x^{jk})}\,\mathrm{d}\epsilon$.

3. The marginal transition function P_{jk} is independent of any initial distribution. And for sets $A_h\in\mathcal{B}^{\infty}(\mathbb{R}^3)$ with $\lim_{h\to0}A_h=\emptyset$ it is for the compound transition probability: $\lim_{h\to0}\limsup_{\xi\to x^{jk}}\int P((\xi,\epsilon),A_h)\,|\epsilon|\,\mathrm{d}\epsilon=0$ for all $x^{jk}\in\mathcal{G}_{jk}$.

4. ϵ has bounded support $\mathcal{G}_0$ and the set $\bar{\mathcal{G}}_{jk}\otimes\mathcal{G}_0$ is small for (X^{jk},ε), where $\mathrm{int}_h\left(\bar{\mathcal{G}}_{jk}\right)=\mathcal{G}_{jk}$.

5. The support of the function m is in $\mathcal{G}$. Its second partial derivatives and are Lipschitz continuous.

Finiteness of the measure π^{ε}_{jk} on $\mathcal{G}_{jk}$ in Assumption 3.1 implies that the asymptotic behavior of the compound process (X^{jk}, ε) is dominated by the β–null structure of the X^{jk} component (see Karlsen, Myklebust, and Tjøstheim (2007), Lemma 6.1.). It is $\pi_{jk}(x) = c\, \pi^{\varepsilon}_{jk}(x)$ with constant $c < \infty$. Thus π^{ε}_{jk} also inherits differentiability properties of π^{jk} from Assumption 2. In Assumption 3.2, the identifying conditional independence criterion is specified. All subsequent assumptions are needed to control the asymptotic behavior of the compound chain. Assumption 3.3 states a local uniform continuity assumption on the transition probability operator P, which allows to control and simplify the variance part in the smoothing as shown in Lemma 5.1. in Karlsen and Tjøstheim (2001). In contrast to the standard or minor Assumptions 3.1 - 3.3, Assumptions 3.4 might appear unusual and artificial. Abstracting from boundedness would require a new way to even prove the standard scalar results in this setting. Smallness, however, is crucial for controlling stochastic terms of the form $f_x(X^{jk}_i, \varepsilon_i) = K_{h,x^{jk}}(X^{jk}_i)\varepsilon_i$ for $x^{jk} \in \mathcal{G}_{jk}$ in the estimators. Under Assumption 1 on the smoothness of the kernel, f is in particular bounded and therefore small and thus special (see Proposition 5.13. in Nummelin (1984)). This implies

$$\sup_{y \in \mathcal{G}_{jk} \times \mathcal{G}_0} \mathbb{E}_y \sum_{i=1}^{\tau} K_{h,x^{jk}}(X^{jk}_i)\varepsilon_i < \infty \quad \text{for all } x^{jk} \in \mathcal{G}_{jk}\ . \tag{4.1}$$

With Assumption 3.5 also $\widetilde{f}_{x^{jk}}(X^{jk}_i) = K_{h,x^{jk}}(X^{jk}_i) m_{jk}(X^{jk}_i)$ is special for each $x^{jk} \in \mathcal{G}_{jk}$ and fulfills (4.1). Compare that in Karlsen, Myklebust, and Tjøstheim (2007) equivalent pointwise conditions were needed to obtain central limit theorems in such a general framework. The support of m must be restricted to $\mathcal{G}$ to control bias terms of the estimator where index sets and observations of different directions mix. This is specific to the nonstationarity in the data and does not appear in stationary SBE.

Remark 2. Note that Assumptions 3 only require a conditional independence condition with respect to invariant measures. Thus short term dependence between residual and covariates is admissible as long as it vanishes asymptotically. This is a much weaker requirement than full independence (see Examples 6.1. and 6.2. in Karlsen, Myklebust, and Tjøstheim (2007) for examples of asymptotically but not fully independent residuals). Thus the estimation problem remains well–posed as long as dependence vanishes asymptotically. In econometrics this is of great importance, as it is contrary to results in the iid case, where any form of endogeneity directly leads to ill–posedness of the problem requiring regularization methods which yield a severely deteriorated small sample behavior (compare Carrasco, Florens, and Renault (2003)). In a special subcase this has been treated in Wang and Phillips (2009b).

If ϵ is ergodic and independent of X, the boundedness and smallness assumption simplify to standard mixing and moment conditions. The subcases with ϵ stationary are of particular interest since they can be regarded as an additive cointegration type model. Assume the following holds for any bivariate marginal process X^{jk} with $j, k = 1, \ldots, d$, $j \neq k$.

ASSUMPTION 3*. 1. X^{jk} and ϵ are independent Harris recurrent Markov chains.

2. ε is ergodic strongly α–mixing with mixing rate satisfying $\sum_l l^{[2/k]\vee 1}\alpha_l < \infty$, $\mu(\varepsilon) = \int \varepsilon \pi_\varepsilon(\varepsilon)\, \mathrm{d}\varepsilon = 0$ and $\int \varepsilon^{p(k+1)} \pi_\varepsilon(\varepsilon)\, \mathrm{d}\varepsilon < \infty$ with $p, k \geq 1$.

3. For sets $A_h \in \mathcal{B}^\infty(\mathbb{R}^2)$ with $\lim_{h\to 0} A_h = \emptyset$ the transition probability of X^{jk} fulfills $\limsup_{\xi\to x} \lim_{h\to 0} P((\xi), A_h) = 0$ for all $x \in \mathcal{G}^{jk}$.

4. The support of the function m is in $\mathcal{G}$. Its second partial derivatives and are Lipschitz continuous.

Remark 3. If all moments on the residual process are finite, it is sufficient if there exists a $\delta > 0$ such that $\sum_l \alpha_l^{1-\delta} < \infty$ for the mixing coefficients.

Note that in general we need the existence of at least the 8th moment in the error term. Though if ε is strictly stationary linear, the moment conditions in Assumption 3* can be relaxed. If ε is strictly stationary linear, it can be written as $\varepsilon_i = \sum_{k=0}^{\infty} a_k e_{i-k}$ with coefficients $\sum_k |a_k| < \infty$ and e strictly stationary with $\mathbb{E}e_0 = 0$, $\mathbb{E}e_0^4 < \infty$, and ϕ–mixing[2] with $\sum_l \phi_l^{1/2} < \infty$. These conditions can replace Assumption 3*.2. They are trivially fulfilled for e_i iid.

For each component function m_j there will be the worst case bivariate nonstationary type dominating the asymptotic behavior through its smallest effective sample size. Therefore denote $\beta^{j+} = \beta^{jk_0} = \min_{k\neq j} \beta^{jk}$ and with this k_0 set $\widehat{L}_{j+} = \widehat{L}_j^{(k_0)}$ for all $j = 1, \ldots, d$. Set $Q_j^{kl} = \beta^{jl}/\beta^{jk}$ for $\beta^{jl} < \beta^{jk}$ and $Q_j^{kl} = -1/2$ for $\beta^{jl} = \beta^{jk}$. Fix $0 < \delta_{j+} < 1$, then $L(n)n^{-\delta_{j+}} \to 0$ with the corresponding slowly varying function for the process X^{jk_0} from (2.2). Hence we get the following closed form expansion.

Theorem 4.1. *Let the model be additive as in* (1.1) *fulfilling the centering condition* (3.7) *and let Assumptions 1-3 hold. Choose a bandwidth sequence such that* $h_{j+} = n^{-\lambda\beta^{j+}}$ *with* $0 < \lambda < \min(1 - \frac{\delta_{j+}+\kappa}{\beta^{j+}}, 1/2 - 1/2Q_j^{kl})$ *for all* l, k *and* $\kappa > 0$ *is arbitrarily small. Then the algorithm* (3.8) *converges with geometric rate and for the estimators* $\widetilde{m}_j(x^j), j = 1, \ldots, d$ *we find*

$$\sqrt{\widehat{L}_{j+}(x^j)h_{j+}}\left(\widetilde{m}_j(x^j) - m_j(x^j) - B_j(x^j)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_{j+}^2(x^j)\frac{\kappa_0^2(x^j)}{\kappa_0(x^j)^2}\right) . \tag{4.2}$$

The bias B_j *consists of two main parts* $B_j(x^j) = B_j^A(x^j) + B_j^B(x^j)$. *The stationary part* B_j^B *coincides in form with the one in the stationary subcase of SBE under the stated bandwidth conditions. The nonstationary part* B_j^A *is specific to GSBE. Both terms vanish with order* h_{j+}^2 *in the interior and with order* h_{j+} *on the boundary. Exact forms are given below. The variance is*

$$\sigma_{j+}^2(x^j) \quad = \quad \int \epsilon^2 \frac{\pi_{j\epsilon}^{k_0}(x^j, \varepsilon)}{\pi_j^{(k_0)}(x^j)}\, d\epsilon \, .$$

Note that the restriction on the bandwidth implies $n^{\beta_{j+}-\delta_{j+}}h_{j+} \gg n^\kappa \to \infty$. As $\kappa > 0$ can be arbitrarily small, this imposes only a mild additional requirement in comparison to marginal local constant estimation in this setting which only needs $n^{\beta_{j+}-\delta_{j+}}h_{j+} \to \infty$. The second restriction guarantees the standard leading bias terms as in the stationary setting. If nonstationary types β^{jl} differ, it ensures that they are far enough apart such that $\beta^{jk} < (1-2\lambda)\beta^{jl}$ wlog for $\beta^{jk} < \beta^{jl}$. In case of violation, the procedure would still be consistent, but leading bias terms were of smaller order than h^2 and of nonstandard form. If all β^{jl} are equal for all l, the restriction is not binding.

[2]See Hall and Heyde (1980), page 277 for an exact definition of ϕ–mixing

As in the stationary case, the deterministic bias B_j^B in Theorem 4.1 consists of three main parts.

$$B_j^B(x^j) \quad = \quad h_{j+}\frac{\kappa_1(x^j)}{\kappa_0(x^j)}m_j'(x^j) + \frac{1}{2}h_{j+}^2\frac{\kappa_2(x^j)}{\kappa_0(x^j)}m_j''(x^j) + \left((I-A)^{-1}\bar{B}\right)_{(j)}(x^j) - b_{j,n}$$

In addition to the marginal Nadaraya–Watson bias for data X_i^j with $i \in I_{jk_0}$ there is the constant shift $b_{j,n}$ from norming and centering and a design density dependent part $\left((I-A)^{-1}\bar{B}\right)_{(j)}(x^j)$. With A the limit of the backfitting operator matrix as with entries as in (3.12), component functions $\bar{b}_j$ for $j \in \{1,\ldots,d\}$ of $\bar{B}(x) = (\bar{b}_1(x^1),\ldots,\bar{b}_d(x^d))^T$ are defined as

$$\bar{b}_j(x^j) \quad = \quad h_{j+}^2\left[\left(b_j + \sum_{k\neq j}\int_{\mathcal{G}_k^{(j)}} b_{jk}(x^k)\frac{\pi_{jk}(x^k)}{\pi_j}\mathrm{d}x^k\right)\right](x^j)\ ,$$

with $b_j(x^j) = \frac{\kappa_2(x^j)}{\kappa_0(x^j)}\left(\frac{m_j'(x^j)}{\pi_j(x^j)}{\pi_j}'(x^j)\right)$ and $b_{jk}(x^{jk}) = \frac{\kappa_2(x^j)}{\kappa_0(x^j)}\left(\frac{m_k'(x^k)}{\pi_{jk}(x^{jk})}\ \frac{\partial\pi_{jk}(x^{jk})}{\partial x^k}\right)$. Most importantly it is $\bar{b}_j = O\left(h_{j+}^2\right)$. Furthermore the centering constant $b_{j,n}$ only affects the level and not the shape of the estimator and is given by $b_{j,n} = \mu_{(j)}\left(\widehat{\Phi}_j\widehat{m}_j\right)$ where the centering operator is defined in (3.10), and it is $h_{j+}^2 b_{j,n} = O(1)$ due to the centering constraint in the algorithm.

The price for nonstationarity in B_j^B occurs through the two dimensional type of nonstationarity driving the rates of order h_{j+}^2 in the interior of $\mathcal{G}_j$ instead of the significantly faster $o(h_j^2)$ as in the standard univariate regression case. Furthermore as generally no full dimensional π exists, the design dependent deterministic bias part resembles its counterpart in the stationary case in form, but lacks its projection interpretation in the empirical $L_{2,\pi}$-norm.

For the stochastic bias it is $B_j^A = ((I-A)^{-1}\bar{B}^A)_j$, where $\bar{B}^A$ has components $\bar{b}_j^A$ of the form

$$\bar{b}_j^A(x^j) \quad = \quad \frac{T^{(j+)\varepsilon}(n)}{\widehat{L}^{j+}(x^j)}\int K_{x^j,h}(X^{j+})\varepsilon\pi_{j\varepsilon}^{(k_0)}\,\mathrm{d}\varepsilon\,\mathrm{d}x^j\ .$$

It vanishes with $o_P(h_{j+}^2)$ under the stated bandwidth assumptions (See Karlsen, Myklebust, and Tjøstheim (2007)). This results in the following Corollary.

Corollary 4.2. *Let the model be additive as in* (1.1) *fulfilling the centering condition* (3.7) *and let Assumptions 1-3 hold. Assume we can choose a bandwidth sequence such that* $h_{j+} = n^{-\lambda\beta^{j+}}$ *with* $1/5 < \lambda < \min(1 - \frac{\delta_{j+}+\kappa}{\beta^{j+}}, 1/2 - 1/2Q_j^{kl})$ *where* $\kappa > 0$ *is arbitrarily small. Then the algorithm* (3.8) *converges with geometric rate and for the estimators* $\widetilde{m}_j(x^j), j = 1,\ldots,d$ *we find*

$$\sqrt{\widehat{L}_{j+}(x^j)h_{j+}}\left(\widetilde{m}_j(x^j) - m_j(x^j)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_{j+}^2(x^j)\frac{\kappa_0^2(x^j)}{\kappa_0(x^j)^2}\right)\ . \tag{4.3}$$

The variance is as in Theorem 4.1.

With the bound from above on the bandwidth, the bias from Theorem 4.1 vanishes. As δ and κ are small such a choice of bandwidth is possible. The following remarks apply to Theorem 4.1 and Corollary 4.2.

Remarks 4. 1. The marginal variance $\sigma_{j+}^2(x^j)$ of the jth additive component is in form exactly the variance of the one–dimensional smoother. Though the rate of convergence is of univariate character

in its form but governed by the worst case bivariate nonstationarity type β^{j+} for each component function.

2. In the case of identical asymptotic order of $\widehat{L}_{j+}$ for all j, the convergence of all component functions in (4.3) holds jointly to a multivariate normal with variance where the only non-zero elements are on the diagonal $(\sigma_{1+}, \ldots, \sigma_{d+})$. This is the case if all pairwise β's coincide - thus all pairs of regressors are of the same type of nonstationarity.

3. The underlying true model can be more general than just additive. For at least pairwise additive models, GSBE yields the best additive approximation (see (3.14)).

4. The results also hold more generally for a model with transformed error term $g_\varepsilon(\varepsilon)$ when replacing ε in Assumption 5.3 by $g_\varepsilon(\varepsilon)$. Then Theorem 4.1 holds with modified asymptotic bias $B_j^A(x^j)$ and variance $\sigma_j(x^j)$ as described in Karlsen, Myklebust, and Tjøstheim (2007).

With stronger independence assumptions on the error term, the stochastic bias B_j^A term in Theorem 4.1 can be omitted. If we enforce the independence assumption between X^j and ϵ, from conditional independence to full independence, we can simplify the boundedness and small set assumptions to more familiar moment conditions. If in addition ε is assumed to be stationary, also the variance is no longer only a second moment with respect to an invariant measure but with respect to the stationary density of ε, hence a "real" variance.

Theorem 4.3. *Let the same set of assumptions as in Theorem 4.1 hold, but replace Assumptions 3 by Assumptions 3*. Choose a bandwidth sequence such that* $h = n^{-\lambda\beta^{j+}}$ *with* $0 < \lambda < \min(1 - \frac{\delta_{j+}+\kappa}{\beta^{j+}}, 1/2 - 1/2Q_j^{kl})$ *where* $\kappa > 0$ *is arbitrarily small. And impose*

$$p > 1 + \frac{\beta^{j+}(3 + \lambda(1 - \frac{1}{k+1})) - 2\omega - 3\delta_{j+} + 2}{2\omega}$$

with $0 < \omega < \kappa_{j+}$ *in Assumption 3*.*
Then the algorithm converges and we get the following asymptotic expansion for the smooth backfitting estimates $(\widetilde{m}_j)_{j=1}^d$

$$\sqrt{\widehat{L}_j^{(k_0)}(x^j)h_{j+}}\left(\widetilde{m}_j(x^j) - m_j(x^j) - B_j^B(x^j)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma_j^2(x^j)\frac{\kappa_0^2(x^j)}{\kappa_0(x^j)^2}\right)$$

with deterministic bias as in Theorem 4.1 and simplified variance

$$\sigma_j^2(x^j) \quad = \quad \sigma_j^2 = \int \varepsilon^2 \pi_\varepsilon(\varepsilon)\, d\varepsilon\,,$$

where π_ε *is the stationary density of* ε*.*

In this setting the stochastic bias B_j^A is zero due to the independence and zero mean assumption.

Observe that in all of the above theorems, the convergence of estimated component functions in GSBE is not jointly - as the scaling of the algorithm and rates of convergence are specific to and can differ for each component. In the special case of one nonstationary regressor and all other regressors

stationary, the nonstationarity influences the rates in all component functions due to choice of effective observations in the estimation. In such an extreme case, tailored methods can improve on feasibility by using the information which regressors are stationary and providing faster rates in these directions. This, however, is technically involved and therefore treated in another paper.

4.2 Discussion

Optimality of the GSBE estimates can be judged against the oracle bound. The infeasible oracle estimator estimates each component function as if all other components were known correctly. It is defined as

$$\ddot{m}_j(x^j) = \frac{1}{d-1}\sum_{k\neq j}\frac{\sum_{i\in I_{jk}} Y_i^* K_{x^j,h}(X_i^j)}{\sum_{i\in I_{jk}} K_{x^j,h}(X_i^j)} \quad \text{with } Y_i^* = Y_i - \sum_{k\neq j} m_k(X_i^k) \tag{4.4}$$

Under standard regularity assumptions as in Karlsen, Myklebust, and Tjøstheim (2007) and Assumptions 2 and 3, this infeasible estimator converges with rate $\sqrt{\widehat{L}_{j+}(x^j)h_{j+}}$ and has variance $\sigma^2_{j+}(x^j)$. These – of univariate form and worst case bivariate type of nonstationarity – are also obtained by GSBE. Due to its local constant form, GSBE contains a design dependent projected part in addition to the oracle bias. We conjecture that a local linear version of GSBE would also reach the oracle bias of a respective local linear benchmark as in the stationary setting. Though as asymptotic results have not even been derived for general nonparametric local linear estimation in this setting in the literature so far, a thorough treatment of this case is beyond the scope of this paper.

If more than just pairs but all higher γ dimensional components of the covariate vector are β–null Harris recurrent, tailored GSBE type methods can improve on projection character and thus on admissible model generality. This was briefly indicated in subsection 3.1. We will shortly illustrate options and implications in the case $\gamma = d$ of a fully β–null Harris recurrent process to highlight the result for GSBE. In this setting, construct all relevant objects according to the full dimensional process analogous to (2.13),(2.14) and replace corresponding objects in the GSBE equation (3.5) by $\widehat{\pi}_j^f$, $\widehat{\pi}_{jk}^f$ and $\widehat{m}_j^f$. Then such adapted GSBE equations reduce in form to the standard smooth backfitting equations (3.3). Therefore projection properties hold as in the stationary case where the algorithm yields the best additive fit without restrictions even if the underlying model is not additive. However, we will see that with this procedure we can reduce the standard stationary type curse of dimensionality, whereas the nonparametric curse of dimensionality remains untouched.

Define the adapted GSBE estimators $\widetilde{m}_j(x^j), j = 1,\ldots,d$ as the iterative solution of the set of equations (3.3) and the normalization (3.2) with the above mentioned modifications. With $\widetilde{m}_0 = \frac{1}{T(n)}\sum_{i\in I} Y_i$ centering can be omitted in the algorithm. Asymptotic properties of the estimators under the most general assumptions on the error are stated in the following theorem.

Theorem 4.4. *Let the model be additive as in* (1.1) *and analogues for* $X_{jk}^{(f)}$ *to Assumptions 1-3 hold. Choose a bandwidth sequence such that* $h = n^{-\lambda\beta_f}$ *with* $0 < \lambda < 1 - \frac{\delta_f+\kappa}{\beta_f}$ *where* $\kappa > 0$ *is arbitrarily small. Then the algorithm* (3.8) *converges with geometric rate and for the estimators* $\widetilde{m}_j(x^j), j = 1,\ldots,d$ *we*

find

$$\sqrt{\widehat{L}_j^f(x^j)h_f}\left(\widetilde{m}_j(x^j) - m_j(x^j) - B_j(x^j)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, (\sigma_j^f)2(x^j)\frac{\kappa_0^2(x^j)}{\kappa_0(x^j)^2}\right) . \tag{4.5}$$

The bias B_j consists of two major parts $B_j(x^j) = B_j^A(x^j) + B_j^B(x^j)$, a stationary part B_j^B which remains and equals the one in the stationary subcase of SBE and a nonstationary part B_j^A. Both terms vanish of order h^2 in the interior and of order h on the boundary. Exact are analogues to the corresponding terms in Theorem 4.1. The variance is

$$(\sigma^f)_j^2(x^j) \quad = \quad \int \epsilon^2 \frac{\pi_{j\epsilon}^f(x^j, \varepsilon)}{\pi_j^f(x^j)}\, d\epsilon \, .$$

Note that the speed of convergence for the smooth backfitting estimator is governed by the effective sample size $\widehat{L}_{x^j,h}^f$ of the full dimensional process which is in general significantly smaller than the rates from GSBE as stated above. The procedure is oracle in rate and variance on the effective observations $X_i \in \mathcal{G}$.

5 Finite Sample Behavior: A Simple Simulation Study

In this section we present simulation results illustrating the finite sample performance of GSBE. In particular we focus on settings where the full dimensional vector of covariates X is no longer recurrent but only pairwise recurrent - thus where general nonparametric estimation might yield inconsistent results. In order to demonstrate that the number of regressors does not affect GSBE, we generate covariates as five dimensional random walks.

In all simulation experiments estimation is repeated 500 times from $n = 1000$ or $n = 10000$ observations. Figure 1 is based on the following model for $i = 1, \ldots, n$

$$\begin{aligned} Y_i &= \sum_{j=1}^{5} m_j(X_i^j) + \epsilon_i \\ X_i &= X_{i-1} + e_i \, , \end{aligned}$$

where $X_0 = (0, 0, 0, 0, 0)^T$ and $m_j(x) = \cos(2\pi(x - 0.5))$ for $j \in \{2, 4\}$ and $m_j(x) = \sin(\pi x)$ for $j \in \{1, 3, 5\}$. The residuals are independent $\varepsilon \sim N(0, \sqrt{0.5})$ and $e \sim N(0, \sigma)$ with $\sigma = ((\sigma_{jk}))_{jk} \in \mathbb{R}^{5\times 5}$. To underline the robustness of the method, we simulate settings with independent random walks as well as cases with correlation, where some off–diagonal elements σ^{kj} are strictly positive. For comparison we also use a local linear version of GSBE which however only provides mild improvements in the finite sample bias (see Table 1) but shows robustness problems in regions of sparse data. The specific model setup is chosen in order to have an easy comparison to the stationary smooth backfitting case, in particular to the extensive simulation study in Nielsen and Sperlich (2005) which focuses on trigonometric relationships. Practically, such models appear in macroeconomic business cycle literature.

In contrast to stationary data, a general β–null Harris recurrent process can cluster in some regions of the space while leaving others almost empty depending on the starting point of the process. And this

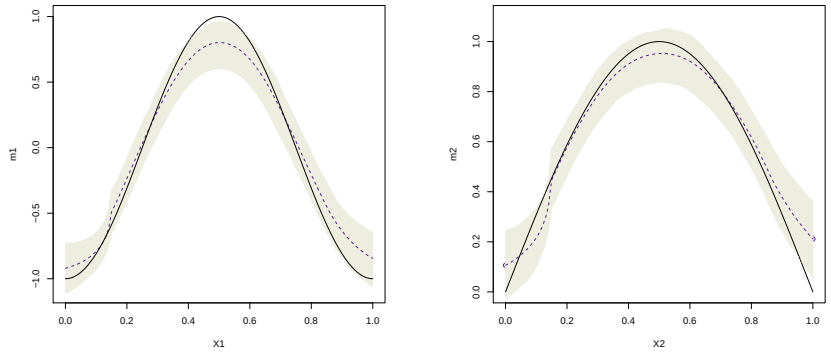

Figure 1: Pointwise median local constant type GSBE fit (violet dashed) in comparison to the respective true function (black solid) for 1000 observations. The shaded areas denote respective interquartile ranges i.e. the 75% and 25% gridpointwise quantiles over all iterations.

phenomenon becomes even more pronounced the higher the dimension of the process. Therefore in finite samples, we can either keep a point x fixed and wait until there are sufficiently many observations in a small neighborhood of x or we can choose a central realization-dependent value of x such as the mode of the sample with additional stochastics entering the problem. Results are, however, quite similar and Figure 2 is representative for both situations.

The issue of a correct finite sample choice of bandwidth is still open, as the theorems only provide asymptotic guidance and β is unknown and hard to estimate from data with 1000 observations (see Remark 3.7. in Karlsen and Tjøstheim (2001)). Here it proved to be useful to select a local bandwidth depending on x via a cross validation criterion for the best componentwise fit. However, rigorous proofs for such a local bandwidth choice require the development of empirical process tools in a Harris recurrent setting. This is quite involved and treated in a separate paper (see also a first attempt in Guerre (2004)). We use $h_j(x^j) \sim \max_k \left(\sum_{i \in I_{jk}} \mathbf{1}_{\mathcal{N}_{x^j}}(X_i^j)\right)^{-1/5}$ for a fixed small neighborhood around x^j. Nonstationary analogues of global cross validation procedures as in Nielsen and Sperlich (2005) are not advisable as they might even in a stationary case induce additional bias of unknown size. We conjecture that the penalized least squares criteria for SBE as in Mammen and Park (2005) can be extended to the general GSBE setting. While simulation results are promising, a theoretical proof for these methods in a β–null Harris recurrent setting requires non-standard higher order expansions which are left for future research.

In the implementation of the algorithm any iteration steps are performed on a fixed grid in each direction. Furthermore without loss of generality the empirical support of the data is transformed to the cuboid $[0,1]^d$. In order to reduce numerical errors in the integrals, $M = 101$ equidistant grid points are chosen. For the algorithm to stop, the following quotient criterion is employed: If

$$\frac{\sum_{i=1}^{M}\left(\widetilde{m}_j^{[r-1]}(x_i^j) - \widetilde{m}_j^{[r]}(x_i^j)\right)^2}{\sum_{i=1}^{M}(\widetilde{m}_j^{[r]}(x_i^j))^2 + 0.0001} < 0.0001 \tag{5.1}$$

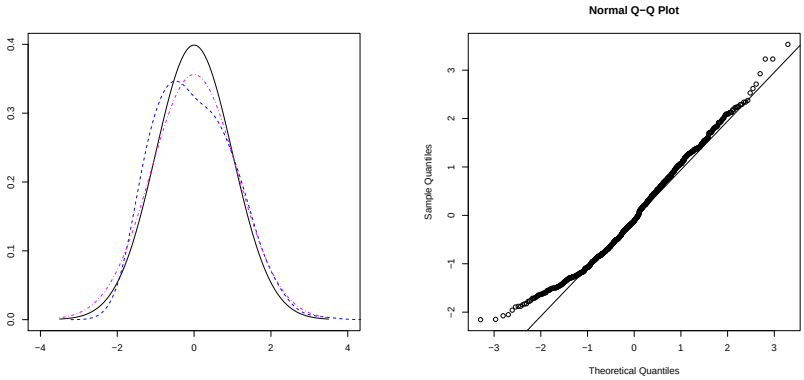

Figure 2: Asymptotic performance of GSBE in component function m_2 using observations in $\mathcal{N}_{x_M,h}$, where x_M is the modal point of the sample: Estimated pdfs of the left hand side of (4.2) for 1000 (dashed blue) and 5000 (magenta dashed dotted) time series observations are compared to the normal pdf (solid black). The corresponding qq-plot for 1000 observations is on the right

is fulfilled for all $j = 1, \ldots, d$ at the M grid points, then end at iteration step r. In the local linear case the algorithm needs on average 20.442 iterations in order to converge while only 15.406 are needed in the local constant case. For comparison, in the stationary convergence is reached after about 6 iterations (see Nielsen and Sperlich (2005)). We compare the fit of different estimators via the median of the integrated square error ISE^k for each additive component.

$$ISE(m_j) = \frac{1}{101}\sum_{l=1}^{101}(m_j(x_l^j) - \widetilde{m}_j(x_l^j))^2 \quad \text{for all } j \in \{1, \ldots, 5\} \ , \tag{5.2}$$

on the grid $0 = x_0 < \ldots < x_l < \ldots < x_{100} = 1$ with $x_l = l \cdot 0.01$, $l = \{0, \ldots, 100\}$.

6 Conclusion

We have introduced a nonparametric estimation procedure, which allows to estimate a regression problem with many potentially nonstationary covariates. Consistent estimators can be obtained even if underlying full-dimensional processes are transient, as long as all possible pairs satisfy a recurrence property. Estimating an additive model allows to circumvent the nonstationary curse of dimensionality as well as the standard ordinary one. Thus obtained rates of convergence and asymptotic variance of GSBE are of univariate form but are governed by the worst case bivariate type of nonstationarity and the corresponding β. In the special case of a stationary residual ϵ, results can be regarded as estimates of additive nonlinear type of cointegration relations.

As in the preceding literature on nonparametric regression for nonstationary covariates, the framework allows for stochastically nonstationary regressors of random walk type but does not cover deterministic trends. Thus the function m might contain a constant m_0, but a deterministic term depending on the

Table 1:

type of fit	underlying data			medianISE on $[0,1]^5$				
	N	σ^{jj}	σ^{kj}	m_1	m_2	m_3	m_4	m_5
Local linear	10,000	1	0	0.012	0.007	0.009	0.007	0.012
Local constant	10,000	1	0	0.026	0.016	0.031	0.019	0.029
Local linear	10,000	1	0.2	0.013	0.009	0.011	0.008	0.012
Local constant	10,000	1	0.2	0.027	0.016	0.031	0.018	0.027
Local linear	1,000	1	0	0.022	0.018	0.018	0.018	0.021
Local constant	1,000	1	0	0.031	0.021	0.034	0.022	0.033
Local linear	1,000	1	0.2	0.027	0.021	0.019	0.020	0.026
Local constant	1,000	1	0.2	0.030	0.017	0.033	0.020	0.033

time parameter is not included in the model. Though extending the model in this direction would introduce challenging problems: properties of estimates for both m and the trend are interrelated and not straightforward to derive even if a parametric form of trend is assumed, and in addition to Harris recurrence properties, appropriate assumptions on the deterministic growth rate of response and regressors are required. This is therefore left for future research.

Though with increasing availability of large data sets, such as e.g. high-frequency data in finance, data on energy consumption, and recent weather and climate data, the method can be a very useful tool in analyzing and testing structural form assumptions. Therefore a clear goal in future research is to develop general statistical testing procedures for cointegration relations with many regressors as done in the case of univariate autoregression in Gao, King, Lu, and Tjøstheim (2009). For implementation of the method in finite samples, however, correct choices of bandwidth in this setting are still an open issue. Furthermore appropriate bootstrap procedures using recurrence times as natural determinants of block sizes are needed. For increased feasibility, we are also working on general semiparametric techniques including cases where some regressors are known to be stationary (Schienle, 2008). This extends (partially) linear models considered in Cai, Li, and Park (2009) or Bandi and Phillips (2007) for linear process structure.

A Proofs

A.1 Preliminaries

In the following, we will make frequently use of uniform consistency results for the density estimator of the invariant measure

$$\widehat{\pi}_{jk}(x^{jk}) = \pi_{jk}(x^{jk}) + o_P\left(n^{-\alpha_1\beta_{jk}+\varepsilon_1}\right) \tag{A.1}$$

$$\widehat{m}_j(x^j) = m_j(x^j) + o_P\left(n^{-\alpha_2\beta_j+\varepsilon_2}\right) \tag{A.2}$$

uniformly in $x^{jk} \in \mathring{\mathcal{G}}^{jk}$ with $\varepsilon_1,\ \varepsilon_2 > 0$ small and in the univariate case for $j = k$, $\alpha_1 = 2/5$, in the bivariate case $j \neq k$, $\alpha_1 = 3/10$, and $\alpha_2 = 2/5$ for optimal bandwidth choice $h = n^{-\beta/5}$. To our knowledge these results are nonstandard and new to the literature for the generality of β–null Harris recurrent processes. They will be derived below in Lemma A.2 and A.3,A.4.

Decompose the one–dimensional pilot smoothers with bivariate type rate of convergence $\widehat{m}_j^{(k)}(x^j)$ into bias $\widehat{m}_j^{(k),B}(x^j)$ and stochastic part $\widehat{m}_j^{(k),A}(x^j)$ as the underlying model (1.1) has an additively separable error term

$$\begin{aligned}
\widehat{m}_j^{(k)}(x^j) &= \frac{\sum_{i\in I_{jk}} K_{h,x^j}(X_i^j)Y_i}{\sum_{i\in I_{jk}} K_{h,x^j}(X_i^j)} \\
&= \left(\frac{\sum_{i\in I_{jk}} K_{h,x^j}(X_i^j)m(X_i)}{\sum_{i\in I_{jk}} K_{h,x^j}(X_i^j)}\right) + \left(\frac{\sum_{i\in I_{jk}} K_{h,x^j}(X_i^j)\epsilon_i}{\sum_{i\in I_{jk}} K_{h,x^j}(X_i^j)}\right) \\
&=: \widehat{m}_j^{(k),B}(x^j) + \widehat{m}_j^{(k),A}(x^j)\ .
\end{aligned} \tag{A.3}$$

As SBE only conducts linear operations on functions, this bias variance separation is preserved in the estimates $(\widetilde{m}_j(x^j))_{j=1}^d$. Thus we have

$$\widetilde{m}_j(x^j) = \widetilde{m}_j^B(x^j) + \widetilde{m}_j^A(x^j)\ , \tag{A.4}$$

where each of the parts $(\widetilde{m}_j^s(x^j))_{j=1}^d$ with superscript $s \in \{A, B\}$ separately solves the defining equations (3.13) for all $j = 1, \ldots, d$

$$\widetilde{m}_j^s(x^j) = \frac{1}{d-1}\mathrm{diag}([I-\Phi]\widehat{m}^s)_j(x^j) + ([A]\widetilde{m}^s)_j(x^j)\ . \tag{A.5}$$

with backfitting projection operator as defined in (3.11).

A.2 Proofs of the Theorems

From the operator backfitting equation (3.13), we can deduce:

$$\begin{aligned}
\widetilde{m} &= (I-\widehat{A})^{-1}\frac{1}{d-1}\mathrm{diag}(\mathbf{1}-\widehat{\Phi})\widehat{m} - (I-\widehat{A})^{-1}(I-\widehat{A})m \\
&= (I-A)^{-1}\left[\frac{1}{d-1}\mathrm{diag}(\mathbf{1}-\widehat{\Phi})\widehat{m} - (I-\widehat{A})m\right] + \\
&\quad + \left((I-\widehat{A})^{-1} - (I-A)^{-1}\right)\left[\frac{1}{d-1}(\mathbf{1}-\widehat{\Phi})\widehat{m} - (I-\widehat{A})m\right]\ .
\end{aligned} \tag{A.6}$$

If we set m_0 as in (3.9), the centering operation with Φ can be omitted. With the uniform results (A.17), it follows from Mammen and Linton (2005) equation (41) that the second summand is negligible for the bias since $\left((I-\widehat{A})^{-1}-(I-A)^{-1}\right)=O_P(h_{jk_0}^2)$ in the interior. Set $\widehat{m}_j^{II}=\frac{1}{d-1}(\mathrm{diag}(\mathbf{1}\widehat{m}))_j$. Focussing on the term in squared bracket it can be shown with the uniform results (A.17) and (A.27) that uniformly over all $x^j\in\mathring{\mathcal{G}}^j$

$$\begin{aligned}\widehat{m}_j^{II}(x^j) &= [(I-\widehat{A})m]_j(x^j)+\left[(I-A)\left(h\frac{\kappa_1}{\kappa_0}m'+\frac{1}{2}h^2\frac{\kappa_2}{\kappa_0}m''\right)\right]_{(j)}(x^j)+\\ &+\left[\frac{1}{2}h^2\frac{\kappa_2}{\kappa_0}\left(b_j+\sum_{k\neq j}\int_{\mathcal{G}_k}b_{jk}(x^k)\frac{\pi_{jk}(x^k)}{\pi_j}\mathrm{d}x^k\right)\right](x^j)+\bar{B}_j^A(x^j)+o_P(n^{-2/5\beta+\varepsilon_1})\end{aligned}\tag{A.7}$$

where h depends on jk_0 and $b_j, b_{jk}, \bar{B}_j^A$ are as specified below Theorem 4.1. Then the bias expression of GSBE is obtained from plugging (A.7) into (A.6). With the uniform result (A.4) and direct analogues of Theorem 2 and 3 in Mammen, Linton, and Nielsen (1999), the asymptotic distribution of GSBE follows from $\widehat{m}_j^{II,A}$ of which the asymptotic distribution corresponds the one of the dominating term $\widehat{m}_j^{(k_0),A}$. Asymptotic normality of $\widehat{m}_j^{(k_0),A}$ follows from Theorem 5.1. in Karlsen, Myklebust, and Tjøstheim (2007).

We are therefore left with the calculation of the explicit bias expansion (A.7) and the proof of explicit uniform results of type (A.1) and (A.2).

Definition A.1. For $g\in L_1(\pi_{jk})$, $h\in L_1(\pi_{j\varepsilon})$ we use the following short hand notation

$$\mu_{jk}(g)=\int_{\mathcal{G}_{jk}}g(u)\pi_{jk}(uv)\,\mathrm{d}u\,\mathrm{d}v\tag{A.8}$$

$$\mu_{(j\varepsilon)}(h\otimes g):=\iint h(u)g(v)\pi_{j\varepsilon}(u,v)\,\mathrm{d}u\,\mathrm{d}v\tag{A.9}$$

For functions with support in

Proof of Expansion (A.7)

With $\widehat{m}_j^{II}$ as before, it is for the bias part

$$\sup_{x^j\in\mathring{\mathcal{G}}^j}\left|\widehat{m}_j^{II,B}(x^j)-\widehat{\nu}_{n,j}(x^j)\right| = o_P(\sum_{k\neq j}h_{jk}^2)\tag{A.10}$$

$$\sup_{x^j\in\partial\mathcal{G}^j}\left|\widehat{m}_j^{II,B}(x^j)-\widehat{\nu}_{n,j}(x^j)\right| = o_P(\sum_{k\neq j}h_{jk})\tag{A.11}$$

where

$$\begin{aligned}\widehat{\nu}_{n,j}(x^j) &= m_j(x^j) + \frac{1}{d-1}\sum_{k\neq j}\sum_{l\neq j}\int_{\mathcal{G}_l}\left(m_l(x^l)\frac{\widehat{\pi}_{jl}(x^j)}{\widehat{\pi}_j^{(k)}(x^j)}\right)\mathrm{d}x^l + \\ &+\frac{1}{d-1}\frac{\kappa_1(x^j)}{\kappa_0(x^j)}\sum_{k\neq j}h_{jk}\left(m_j'(x^j)+\sum_{l\in\mathcal{L}_{jk}}\int_{\mathcal{G}_l}\left(m_l'(x^l)\frac{\widehat{\pi}_{jl}(x^{jl})}{\widehat{\pi}_j^{(k)}(x^j)}\right)\mathrm{d}x^l\right)+ \\ &+\frac{1}{d-1}\frac{\kappa_2(x^j)}{\kappa_0(x^j)}\sum_{k\neq j}h_{jk}^2\left(\frac{1}{2}m_j''(x^j)+m_j'(x^j)\frac{\pi_j{}'(x^j)}{\pi_j(x^j)}+\right. \\ &\left.+\sum_{l\in\mathcal{L}_{jk}}\int_{\mathcal{G}_l}\left(m_l'(x^l)\frac{\partial\pi_{jl}(x^{jl})}{\partial x^l\pi_{jl}(x^{jl})}+\frac{1}{2}m_l''(x^l)\right)\frac{\pi_{jl}(x^{jl})}{\pi_j(x^j)}\mathrm{d}x^l\right)\ .\end{aligned}$$

Furthermore the stochastic part vanishes according to

$$\sup_{x^j\in\mathcal{G}_j}\left|\widehat{m}_j^{II,A}(x^j)-\frac{1}{d-1}\sum_{k\neq j}\mu_{(j\varepsilon)}(K_{x^j,h}\otimes\mathrm{id}_\varepsilon)\frac{T_{jk\varepsilon}(n)}{\widehat{L}_j^{(k)}(x^j)}\right| = o_P\left(\sum_{k\neq j}\frac{1}{\sqrt{n^{\beta^{jk}-\delta}h_{jk}}}\right)\ . \quad (A.12)$$

Combining the expansions for bias part (A.10) and (A.11) with the result for the stochastic part (A.12) yields (A.7) for appropriate bandwidth choices as in all theorems.

Proof. The statement (A.12) for the stochastic part is a direct consequence from the uniform result (A.4), which is shown below.

For the bias part, decompose $\left(\mathrm{diag}(\mathbf{1}\widehat{m}^B)\right)_j$ in the following way

$$\begin{aligned}\left(\mathrm{diag}(\mathbf{1}\widehat{m}_j^B)\right)(x^j) &= \sum_{k\neq j}\widehat{m}_j^{(k),B}(x^j)=\sum_{k\neq j}\frac{1}{T^{jk}(n)}\sum_{i\in I_{jk}}\frac{K_{h_{jk},x^j}(X_i^j)m(X_i)}{\widehat{\pi}_j^{(k)}(x^j)} \\ &= \sum_{k\neq j}\frac{1}{T^{jk}(n)}\sum_{i\in I_{jk}}\frac{K_{h_{jk},x^j}(X_i^j)\left(m_0+\sum_{l=1}^d m_l(X_i^l)\right)}{\widehat{\pi}_j^{(k)}(x^j)} \qquad (A.13)\end{aligned}$$

Convergence holds due to the quotient limit theorem (see e.g. Meyn and Tweedie (1993)). Expand (A.13) for each summand separately. In the numerator of (A.13) distinguish between three cases $l=j$ or $l=k\neq j$ and $l\neq(j\vee k)$ in the summands $K_{h_{jk},x^j}m_l$. We will see that the last case has some nonstationary peculiarities. For $l=j$ it is with (A.27) and standard kernel calculations

$$\begin{aligned}&\sum_{k\neq j}\frac{1}{T^{jk}(n)}\sum_{i\in I_{jk}}\frac{K_{h_{jk},x^j}(X_i^j)m_j(X_i^j)}{\widehat{\pi}_j^{(k)}(x^j)} \\ &= \sum_{k\neq j}m_j(x^j)+\frac{\mu_{jk}\left(K_{h_{jk},x^j}(\cdot)m_j(\cdot)\right)-m_j(x^j)\mu_{jk}\left(K_{h_{jk},x^j}(\cdot)\right)}{\mu_{jk}\left(K_{h_{jk},x^j}(\cdot)\right)}+R_{n_{jk},jk} \\ &= \sum_{k\neq j}m_j(x^j)+h_{jk}\frac{\kappa_1(x^j)}{\kappa_0(x^j)}m_j'(x^j)+h_{jk}^2\frac{\kappa_2(x^j)}{\kappa_0(x^j)}\left(\frac{m_j'(x^j)}{\pi_j(x^j)}\pi_j'(x^j)+\frac{1}{2}m_j''(x^j)\right) \\ &\qquad +R_{n_{jk},jk}+o_P(h_{jk}^2) \\ &= (d-1)m_j(x^j)+\sum_{k\neq j}h_{jk}\frac{\kappa_1(x^j)}{\kappa_0(x^j)}m_j'(x^j)+h_{jk}^2\frac{\kappa_2(x^j)}{\kappa_0(x^j)}\left(\frac{m_j'(x^j)}{\pi_j(x^j)}\pi_j'(x^j)+\frac{1}{2}m_j''(x^j)\right)+o_P(h_{jk}^2) \\ &= (d-1)\left(m_j(x^j)+h_{j+}\frac{\kappa_1(x^j)}{\kappa_0(x^j)}m_j'(x^j)+h_{j+}^2\frac{\kappa_2(x^j)}{\kappa_0(x^j)}\left(\frac{m_j'(x^j)}{\pi_j(x^j)}\pi_j'(x^j)+\frac{1}{2}m_j''(x^j)\right)+o_P(h_{j+}^2)\right)\end{aligned}$$

The last equation is true since for

$$\begin{aligned}
&\sup_{x^j\in\mathcal{G}^j}\left|R_{n_{jk},jk}(x^j)\right| \\
&= \sup_{x^j\in\mathcal{G}^j}\left|\frac{1}{T^{jk}(n)}\sum_{i\in I_{jk}}\frac{K_{h_{jk},x^j}(X_i^j)m_j(X_i^j)}{\widehat{\pi}_j^{(k)}(x^j)}-\frac{\mu_{jk}\left(K_{h_{jk},x^j}(\cdot)m_j(\cdot)\right)}{\mu_{jk}\left(K_{h_{jk},x^j}(\cdot)\right)}\right| \\
&= \sup_{x^j\in\mathcal{G}^j}\left|\frac{1}{T^{jk}(n)}\left[\sum_{i\in I_{jk}}\frac{K_{h_{jk},x^j}(X_i^j)m_j(X_i^j)-\mu(K_{h_{jk},x^j}m_j)}{\widehat{\pi}_j^{(k)}(x^j)}\right]+\right. \\
&\quad\left.+\frac{\mu(K_{h_{jk},x^j}m_j)}{\widehat{\pi}_j^{(k)}(x^j)}-\frac{\mu_{jk}\left(K_{h_{jk},x^j}(\cdot)m_j(\cdot)\right)}{\mu_{jk}\left(K_{h_{jk},x^j}(\cdot)\right)}\right| = o_P\left(\frac{1}{\sqrt{h_{jk}n^{\beta_{jk}-\delta_{jk}}}}\right)=o_P(h_{jk}^2)\,.
\end{aligned}$$

The details of this follow exactly from the proof of expansion (A.2) in Lemma A.2 for the deterministic part. For $l=k\neq j$ standard kernel calculations lead to

$$\begin{aligned}
&\sum_{k\neq j}\frac{1}{T^{jk}(n)}\sum_{i\in I_{jk}}\frac{K_{h_{jk},x^j}(X_i^j)m_k(X_i^k)}{\widehat{\pi}_j^{(k)}(x^j)} \\
&= \sum_{k\neq j}\frac{1}{T^{jk}(n)}\sum_{i\in I_{jk}}\int_{\mathcal{G}_k}\frac{K_{h_{jk},x^j}(X_i^j)K_{h_{jk},x^k}(X_i^k)m_k(X_i^k)}{\widehat{\pi}_j^{(k)}(x^j)}\mathrm{d}x^k \\
&= \sum_{k\neq j}\sum_{i\in I_{jk}}\int_{\mathcal{G}_k}\frac{K_{h_{jk},x^j}(X_i^j)K_{h_{jk},x^k}(X_i^k)}{T^{jk}(n)\widehat{\pi}_j^{(k)}(x^j)}\left(m_k(x^k)+\right. \\
&\qquad\left.+m_k'(x^k)(X_i^k-x^k)+\frac{1}{2}m_k''(x^k)(X_i^k-x^k)^2\right)\mathrm{d}x^k+o_P(h_{jk}^2) \\
&= \sum_{k\neq j}\left[\int_{\mathcal{G}_k}\frac{\widehat{\pi}_{jk}(x^{jk})}{\widehat{\pi}_j^{(k)}(x^j)}m_k(x^k)\mathrm{d}x^k+h_{jk}\int_{\mathcal{G}_k}\frac{\widehat{\pi}_{jk}(x^{jk})}{\widehat{\pi}_j^{(k)}(x^j)}\frac{\kappa_1(x^j)}{\kappa_0(x^j)}m_k'(x^k)\mathrm{d}x^k+\right. \\
&\quad\left.+h_{jk}^2\frac{\kappa_2(x^j)}{\kappa_0(x^j)}\int_{\mathcal{G}_k}\left(\frac{\partial\pi_{jk}(x^{jk})}{\pi_{jk}(x^{jk})\partial x^k}m_k'(x^k)+\frac{1}{2}m_k''(x^k)\right)\frac{\pi_{jk}(x^{jk})}{\pi_j(x^j)}\mathrm{d}x^k\right] \\
&\qquad+R_{n_{jk},jk}(x^j)+o_P(h_{jk}^2)
\end{aligned}$$

For the second to last equation, standard kernel arguments are applied together with (A.17) and Lemma A.3. Exact details follow from Mammen, Linton, and Nielsen (1999) equations (118)-(122). It involves particular showing uniform convergence of order h_{jk}^2 of the following expressions against their respective means,

$$t_j^l(x^j)=K_{h_{jk},x^j}(X_i^j)\int_{\mathcal{G}_k}K_{h_{jk},x^k}(X_i^k)(X_i^k-x^k)^l m_k^{(l)}(x^k)\,\mathrm{d}x^k$$

with $l\in\{1,2\}$, which can be expanded into the terms above. This is achieved along the lines of Lemma A.3. The last equation is true since $\sup_{x^j\in\mathcal{G}_j}\left|R_{n,j}(x^j)\right|=o_P(h_{jk}^2)$. This is shown as before.

For $l\neq(j\vee k)$ the fact that we might use different data in different directions complicates the expansion. We study

$$\sum_{k\neq j}\frac{1}{T^{jk}(n)}\sum_{i\in I_{jk}}\frac{K_{h_{jk},x^j}(X_i^j)m_l(X_i^l)}{\widehat{\pi}_j^{(k)}(x^j)}\,. \tag{A.14}$$

For the index set I_{jk}, data of the marginal X^l might also be found outside $\mathcal{G}_l$. Though these points have no effect in the overall expansion since $m_l(z) = 0$ for $z \in \mathbb{R}\backslash\mathcal{G}_l$ under Assumption 3.5. or 3*.5. Thus expand the λ_{jk} non-zero terms in the numerator of (A.14) if $l \in \mathcal{L}_{jk}$ as defined in (2.16)

$$\begin{aligned}
&\sum_{k\neq j}\frac{1}{T^{jk}(n)}\sum_{i\in I_{jk}}\frac{K_{h_{jk},x^j}(X_i^j)m_l(X_i^l)}{\widehat{\pi}_j^{(k)}(x^j)}\\
&= \sum_{k\neq j}\frac{1}{T^{jk}(n)}\sum_{i\in I_{jk}}\int_{\mathcal{G}_l}\frac{K_{h_{jk},x^j}(X_i^j)K_{h_{jl},x^l}(X_i^l)m_l(X_i^l)}{\widehat{\pi}_j^{(k)}(x^j)}\mathrm{d}x^l\\
&= \sum_{k\neq j}\left[\int_{\mathcal{G}_l}\frac{\widehat{\pi}_{jl}^{(k)}(x^{jl})}{\widehat{\pi}_j^{(k)}(x^j)}m_l(x^l)\mathrm{d}x^l + h_{jk}\int_{\mathcal{G}_l}\frac{\widehat{\pi}_{jl}^{(k)}(x^{jl})}{\widehat{\pi}_j^{(k)}(x^j)}\frac{\kappa_1(x^j)}{\kappa_0(x^j)}m_l'(x^l)\mathrm{d}x^l\ +\right.\\
&\quad + h_{jk}^2\frac{\kappa_2(x^j)}{\kappa_0(x^j)}\int_{\mathcal{G}_l}\left(\frac{\partial\pi_{jl}(x^{jl})}{\pi_j(x^{jl})\partial x^l}m_l'(x^l) + \frac{1}{2}m_l''(x^l)\right)\frac{\pi_{jl}(x^{jl})}{\pi_j(x^j)}\mathrm{d}x^l\Big]\\
&\quad + R_{n,jk}(x^j) + o_P(h_{jk}^2)\ .
\end{aligned}$$

Recall that $\mathcal{L}_{jk}$ collects all summand in (A.14) where numerator and denominator converge almost surely to a nonzero limit. For all other l the last term in the above expansion vanishes as it is of order $h_{jk}^2 \cdot \widehat{L}^{jl}(x^j)/\widehat{L}^{jk}(x^j)$ smaller than h_{jk}^2.

In total adding up with $o_P(\sum_{k\neq j} h_{jk}^2) = o_P(h_{j+}^2)$, claims (A.10) and (A.11) have been proven which yield the expansion terms in (A.7) □

A.3 Proof of the Uniform Results

A.3.1 Proof of (A.1)

To ease notation, indices and superscripts indicating components to be marginal j or jk specific will be generally omitted. We write id_ε for the identity on the support of ε, i.e., $\mathrm{id}_\varepsilon(u) = u$ for $u \in \mathcal{G}_0$. Assume the split chain component U is defined as in (2.3). We need the following moment bounds on $U(K_{x,h})$

Lemma A.1. *Let Assumptions 1-2 hold. Then it is $\mu(|K_{x,h}|) = \mathbb{E}U(|K_{x,h}|) = \pi(|K_{x,h}|) = \mu + o(1)$ with $-\infty < \mu < \infty$ and*

$$\mathbb{E}U^q(|K_{x,h}|) \leq C_1 h^{-(q-1)d} \quad \textit{for } q > 1 \tag{A.15}$$

where $C_1 < \infty$. With Assumption 3 or Assumption 3 in addition it is*

$$\mathbb{E}U^p(|K_{x,h} \otimes id_\varepsilon|) \leq C_2 h^{-(p-\frac{1}{\nu})d} \tag{A.16}$$

with $C_2 < \infty$ and $\nu = 1$ and $p > 1$ arbitrary for Assumption 3 and $\nu = k+1$ and p as in Assumption 3 under the latter. All relations also hold uniformly over x in compact sets.*

Proof. For the proof of the pointwise bounds see Lemma 5.1. and 5.2. in Karlsen and Tjøstheim (2001) for (A.15), for (A.16) see the end of the proof of Theorem 6.1. and Theorem A.1. in Karlsen, Myklebust, and Tjøstheim (2007) under dependence and smallness and Theorem 5.3. in Karlsen, Myklebust, and

Tjøstheim (2007) under independence and mixing conditions. The uniform results are straightforward extensions and essentially follow since Assumption 1(ii) implies that $\mathcal{S}_{x,h} = x \oplus h\mathcal{S}$ is small and any for the sup relevant expression in the proof of Lemma 5.1. and 5.2. in Karlsen and Tjøstheim (2001) can thus be bounded by

$$\sup_{v \in \mathcal{S}_{x,h}} \mathbb{E}_v \sum_{i=0}^{\tau} \mathbf{1}_{\mathcal{S}_{x,h}}(X_i) \leq \sup_{v \in \mathcal{S}_{x,h}} \pi_s(v) \, |\mathcal{S}_{x,h}| \leq C \, .$$

□

Lemma A.2. *Let X be β–null Harris recurrent with continuously differentiable invariant density π. Let Assumptions 1 hold and choose a bandwidth $h = n^{-\lambda\beta/d}$ with $0 < \lambda < 1 - \frac{\delta+\kappa}{\beta}$ where $\kappa > 0$ arbitrarily small. Then*

$$\sup_{x \in \hat{\mathcal{G}}_h} |\widehat{\pi}(x) - \pi(x)| = o_P\left(h^2 + \frac{1}{\sqrt{n^{\beta-\delta}h^d}}\right) \tag{A.17}$$

On the boundary, the rate of bias is only of order h instead of h^2.

Note that under the stated conditions it is $n^{\beta-\delta}h^d = n^{\kappa} \to \infty$ with $\kappa > 0$.

Proof. It is sufficient to show that for $c_n = h^2 + \sqrt{n^{-\beta+\delta}h^{-d}}$ or on the boundary $c_n' = h + \sqrt{n^{-\beta+\delta}h^{-d}}$ respectively, and for all $\eta, \eta' > 0$ there exist constants $c, c' > 0$ such that

$$\sup_n \mathbb{P}\left(\sup_{x \in \mathring{\mathcal{G}}} |\widehat{\pi}(x) - \pi(x)| \geq c \cdot c_n\right) = \eta \tag{A.18}$$

and with c' and η' on the boundary. In fact we will even show almost sure convergence. To shorten notation we write c_n instead of $c \cdot c_n$ for $c \in \mathbb{R}$. In the following C is an arbitrary constant which might vary from line to line.

Split up into variance and bias part. For the interior $\mathring{\mathcal{G}}_h$ it is:

$$\begin{aligned}
&\mathbb{P}\left(\sup_{x \in \mathring{\hat{\mathcal{G}}}_h} |\widehat{\pi}(x) - \pi(x)| \geq c_n\right) \\
&\leq \mathbb{P}\left(\sup_{x \in \hat{\mathcal{G}}_h} |\widehat{\pi}(x) - \mu(K_{x,h})| + \sup_{x \in \hat{\mathcal{G}}_h} |\mu(K_{x,h}) - \pi(x)| \geq c_n\right) \\
&\leq \mathbb{P}\left(\sup_{x \in \hat{\mathcal{G}}_h} |\widehat{\pi}(x) - \mu(K_{x,h})| \geq \frac{c_n}{2}\right) + \mathbb{P}\left(\sup_{x \in \mathring{\hat{\mathcal{G}}}_h} |\mu(K_{x,h}) - \pi(x)| \geq \frac{c_n}{2}\right) \\
&= S_1^i + S_2^i \, ,
\end{aligned}$$

Since $\mathcal{G} \subset \mathbb{R}^d$ is bounded, for the $C_1 h$-ring-boundary $\partial\mathcal{G}_h$ we get:

$$\begin{aligned}
&\mathbb{P}\left(\sup_{x \in \partial\mathcal{G}_h} \left|\widehat{\pi}(x) - \pi(x) \int_{\mathcal{G}} K_{h,x}(u)\mathrm{d}u\right| \geq c_n\right) \\
&\leq \mathbb{P}\left(\sup_{x \in \partial\mathcal{G}_h} |\widehat{\pi}(x) - \mu(K_{x,h})| \geq \frac{c_n}{2}\right) + \mathbb{P}\left(\sup_{x \in \partial\mathcal{G}_h} \left|\mu(K_{x,h}) - \pi(x) \int_{\mathcal{G}} K_{h,x}(u)\mathrm{d}u\right| \geq \frac{c_n}{2}\right) \\
&= S_1^b + S_2^b \, ,
\end{aligned}$$

For the bias parts S_2^i and S_2^b, standard analysis with the usual kernel arguments carries over. Since we have for x in the interior $\mathring{\mathcal{G}}_{C_1 h}$ that $\int_{\mathcal{G}} K_{h,x}(u)\mathrm{d}u = 1$, we can treat S_2^i and S_2^b together:

$$\begin{aligned}
\mu(K_{x,h}) - \pi(x)\int_{\mathcal{G}} K_{h,x}(u)\mathrm{d}u &= \int_{B_x(C_1 h)\cap\mathcal{G}} (\pi(x+hu)-\pi(x))K(u)\mathrm{d}u \\
&= (\pi'(x))\, h \int_{B_x(C_1 h)\cap\mathcal{G}} uK(u)\mathrm{d}u + O(h^2) \\
&= \begin{cases} O(h^2) & \text{for } x \in \mathring{\mathcal{G}}_h \\ O(h) & \text{for } x \in \partial\mathcal{G}_h \end{cases},
\end{aligned}$$

since for $x \in \partial\mathcal{G}_h$ the ball $B_x(C_1 h)$ is not entirely in $\mathcal{G}$. Thus the with symmetry of the kernel the integral is not zero as in the case for x in the interior.

Now treat the stochastic term $S_1 = \mathbb{P}\left(\sup_{x\in\mathcal{G}} |\widehat{\pi}(x) - \mu(K_{x,h})| \geq \frac{c_n}{2}\right)$. Here we do not have to distinguish between cases of x on the boundary or not. As $\mathcal{G}$ is compact, there exists a cover of $l(n)$ open balls $L_1, \ldots, L_k, \ldots, L_{l(n)}$ with radius $\frac{c_1}{l(n)^{1/d}}$ for an appropriate constant c_1 and with centers in x_k and $\bigcup_{k=1}^{l(n)} L_k \supseteq \mathcal{G}$. Set

$$l(n) = \sqrt{n^{\beta d(1+\lambda d)-\delta}};. \tag{A.19}$$

The maximal distance attainable between elements inside one of the balls is the diameter

$$\max_{a,b\in L_k} \|a-b\| \leq \frac{2c_1}{l(n)^{1/d}} = \frac{c}{l(n)^{1/d}} \text{ for all } k \in \{1,\ldots,l(n)\}\ . \tag{A.20}$$

Then

$$\begin{aligned}
&\mathbb{P}\left(\sup_{x\in\mathcal{G}} |\widehat{\pi}(x) - \mu(K_{x,h})| \geq \frac{c_n}{2}\right) \\
&= \mathbb{P}\left(\max_{1\leq k\leq l(n)} \sup_{x\in\mathcal{G}\cap L_k} |\widehat{\pi}(x) - \mu(K_{x,h})| \geq \frac{c_n}{2}\right) \\
&\leq \mathbb{P}\left(\max_{1\leq k\leq l(n)} \sup_{x\in\mathcal{G}\cap L_k} |\widehat{\pi}(x) - \widehat{\pi}(x_k)| \geq \frac{c_n}{6}\right) + \mathbb{P}\left(\max_{1\leq k\leq l(n)} |\widehat{\pi}(x_k) - \mu(K_{x_k,h})| \geq \frac{c_n}{6}\right) \\
&\quad + \mathbb{P}\left(\max_{1\leq k\leq l(n)} \sup_{x\in\mathcal{G}\cap L_k} |\mu(K_{x_k,h}) - \mu(K_{x,h})| \geq \frac{c_n}{6}\right) \\
&= Q_1 + Q_2 + Q_3\ .
\end{aligned}$$

The first and the third term, Q_1 and Q_3, are easy to handle and therefore treated first. Look at Q_1:

$$\begin{aligned}
\sup_{x\in\mathcal{G}\cap L_k} |\widehat{\pi}(x) - \widehat{\pi}(x_k)| &= \frac{1}{T(n)} \sup_{x\in\mathcal{G}\cap L_k} \left|\sum_{i=1}^{n} (K_{h,x}(X_i) - K_{h,x_k}(X_i))\right| \\
&\leq \sup_{x\in\mathcal{G}\cap L_k} \frac{\widetilde{L}}{h_n^{d+1}} \|x - x_k\| \quad \mathbb{P}-a.s. \\
&\leq \frac{\widetilde{L}c_1}{h_n^{d+1} l(n)^{1/d}} \quad \mathbb{P}-a.s.\ .
\end{aligned}$$

The first $\mathbb{P}-a.s$ relation is a consequence of the quotient limit theorem (see, e.g. Meyn and Tweedie (1993)), while the inequalities thereafter follow directly from (A.20) and the Lipschitz assumption on the kernel in Assumption 1.

Since the integral operator and everything inside is continuous, obviously we also get $\max_{1\le k\le l(n)} \sup_{x\in\mathcal{G}\cap L_k} |\mu(K_{x_k,h}) - \mu(K_{x,h})| = O\left(\frac{1}{h_n^{d+1} l(n)^{1/d}}\right)$ a.s.. Thus when imposing $c_n = O\left(h_n^{d+1} l(n)^{1/d}\right)$, then Q_1 and Q_3 are $o_P(1)$.

Q_2, the second term, however, needs some extra considerations: On the grid of the x_k-balls we can simplify in the standard way as

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le k\le l(n)} |\widehat{\pi}(x_k) - \mu(K_{x_k,h})| \ge \frac{c_n}{6}\right) \\
&= \mathbb{P}\left(\max_{1\le k\le l(n)} \left|\frac{1}{T(n)}\sum_{i=1}^{n} K_{x_k,h} - \mu(K_{x_k,h})\right| \ge \frac{c_n}{6}\right) \\
&= \mathbb{P}\left(\max_{1\le k\le l(n)} \left|\frac{1}{T(n)}\left(U_{0,x_k,h} + \sum_{j=1}^{T(n)} W_{j,x_k,h} + U_{n,x_k,h}\right)\right| \ge \frac{c_n}{6}\right) ,
\end{aligned}$$

where the sum is rewritten in terms of the centered split chain components $W_{j,x,h} = U_{j,x,h} - \mu(K_{x,h})$ where $U_{j,x,h}$ is the j–th component of the split chain of $K_{x,h}$ as defined in (2.4). As parts of a split chain all $W_{j,x,h}$ are iid $W_{x,h}$ for a given $x \in \mathcal{G}$. And obviously from the definition it is $\mathbb{E}(W_{x,h}) = 0$. It is easy to show that $\max_{1\le k\le l(n)} \frac{|U_{0,x_k,h}|}{T(n)} = o(1)$ a.s. and $\max_{1\le k\le l(n)} \frac{|U_{n,x_k,h}|}{T(n)} = o(1)$ a.s. (see proof of Theorem 5.1. in Karlsen and Tjøstheim (2001)).
Therefore it suffices to look at $\mathbb{P}\left(\max_{1\le k\le l(n)} \left|\frac{1}{T(n)}\sum_{j=1}^{T(n)} W_{j,x,h}\right| \ge c_n'\right)$. Since c_n' differs from c_n only by a constant, we continue notation with c_n. As the norming $T(n)$ is stochastic and not independent of $W_{x,h}$, use a truncation argument. There exist $\delta_n^{(1)}, \delta_n^{(2)}$ such that

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le k\le l(n)} \frac{1}{T(n)}\left|\sum_{j=1}^{T(n)} W_{j,x_k,h}\right| \ge c_n\right) \\
&\qquad \le \mathbb{P}\left(\max_{1\le k\le l(n)} \frac{1}{T(n)}\left|\sum_{j=1}^{T(n)} W_{j,x_k,h}\right| \ge c_n, \delta_n^{(1)} \le T(n) \le \delta_n^{(2)}\right) + o_P(\nu_n) , \qquad \text{(A.21)}
\end{aligned}$$

with $\nu_n \to 0$ negligible. Set $\delta_n^{(1)} = n^{\beta-\alpha}$ and $\delta_n^{(2)} = n^{\beta+\alpha}$ with $0 < \delta \le \alpha \ll 1$. Then inequality (A.21) follows since

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le k\le l(n)} \frac{1}{T(n)}\left|\sum_{j=1}^{T(n)} W_{j,x_k,h}\right| \ge c_n, T(n) \le \delta_n^{(1)}\right) \le \mathbb{P}\left(T(n) \le n^{\beta-\alpha}\right) \\
&\qquad \le 1 - \frac{(1-n^{-\alpha})n^{\beta}}{(1+n^{-\alpha})n^{\beta}+1} = \frac{2n^{-\alpha}n^{\beta}+1}{(1+n^{-\alpha})n^{\beta}+1} = O(n^{-\alpha})
\end{aligned}$$

where in the second line we use the first assertion of the proof of Theorem 2.1 in Chen (2000). And together with

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le k\le l(n)} \frac{1}{T(n)}\left|\sum_{j=1}^{T(n)} W_{j,x_k,h}\right| \ge c_n, T(n) \ge \delta_n^{(2)}\right) \le \mathbb{P}\left(T(n) \le n^{\beta+\alpha}\right) \\
&\qquad \le \frac{T(n)^m}{(L_s(n)n^{\beta})^m} \cdot \frac{(L_s(n)n^{\beta})^m}{(n^{\beta+\alpha})^m} = O\left((L_s(n))^m n^{-\alpha m}\right)
\end{aligned}$$

we obtain (A.21). In order to obtain the last line we use Markov inequality for $m > 1$ and that the first quotient factor is bounded as shown in Karlsen and Tjøstheim (2001) Lemma 3.3..

Continuing from (A.21) with $\alpha > 0$ it is on the grid

$$\begin{aligned}
&\mathbb{P}\left(\max_{1\le k\le l(n)} \frac{1}{T(n)}\left|\sum_{j=1}^{T(n)} W_{j,x_k,h}\right| \ge c_n, \delta_n^{(1)} \le T(n) \le \delta_n^{(2)}\right) \\
&\quad\le \quad l(n)\cdot \sup_{x\in\mathcal{G}} \mathbb{P}\left(\frac{1}{T(n)}\left|\sum_{j=1}^{T(n)} W_{j,x,h}\right| \ge c_n, \delta_n^{(1)} \le T(n) \le \delta_n^{(2)}\right) \ . \qquad \text{(A.22)}
\end{aligned}$$

For truncating in the elements of the sum, define:

$$\begin{aligned}
\xi_{k,x,h} &= U_{k,x,h}\mathbf{1}_{|U_{k,x,h}|\le R_n} - \mathbb{E}(U_{k,x,h}\mathbf{1}_{|U_{k,x,h}|\le R_n}) \\
\eta_{k,x,h} &= U_{k,x,h}\mathbf{1}_{|U_{k,x,h}|> R_n} - \mathbb{E}(U_{k,x,h}\mathbf{1}_{|U_{k,x,h}|> R_n})
\end{aligned}$$

with $h^{-d} \ll R_n \ll n^{\beta-\alpha} \ll n^{\beta-\delta}$. These restrictions result from (A.24) and (A.25) below

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{T(n)}\left|\sum_{j=1}^{T(n)} W_{j,x,h}\right| \ge c_n, \delta_n^{(1)} \le T(n) \le \delta_n^{(2)}\right) \\
&\quad\le \quad \mathbb{P}\left(\frac{1}{T(n)}\left|\sum_{j=1}^{T(n)} \xi_{j,x,h}\right| \ge \frac{c_n}{2}, \delta_n^{(1)} \le T(n) \le \delta_n^{(2)}\right) \\
&\qquad +\mathbb{P}\left(\frac{1}{T(n)}\left|\sum_{j=1}^{T(n)} \eta_{j,x,h}\right| \ge \frac{c_n}{2}, \delta_n^{(1)} \le T(n) \le \delta_n^{(2)}\right) \qquad \text{(A.23)}
\end{aligned}$$

Treat the two terms separately and continue with c_n.

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{T(n)}\left|\sum_{j=1}^{T(n)} \xi_{j,x,h}\right| \ge c_n, \delta_n^{(1)} \le T(n) \le \delta_n^{(2)}\right) &\le \sum_{k=\lfloor \delta_n^{(1)}\rfloor}^{\lceil \delta_n^{(2)}\rceil} \mathbb{P}\left(\frac{1}{k}\left|\sum_{j=1}^{k} \xi_{j,x,h}\right| > c_n\right) \\
&\le C \sum_{k=\lfloor \delta_n^{(1)}\rfloor}^{\lceil \delta_n^{(2)}\rceil} 2\exp\left(-\frac{2c_n^2 k}{R_n^2}\right) \le C\exp\left(-\frac{c_n^2 \lfloor \delta_n^{(1)}\rfloor}{R_n^2}\right) \ , \qquad \text{(A.24)}
\end{aligned}$$

where $\lfloor z\rfloor$ is the largest possible integer smaller or equal z and $\lceil z\rceil$ is the smallest possible integer greater or equal z. We obtain (A.24) with the standard Hoeffding inequality for iid bounded observations.

For the second term denote the event
$A_i = \left\{|U_{j,x,h}| > R_n \text{ for } j \in J = \{j_1, \ldots, j_i\} \text{ and } \delta_n^{(1)} \le T(n) \le \delta_n^{(2)}\right\} \bigcap C_i = B_i \bigcap C_i$ with $C_i = \{\sum_J |U_{j,x,h}| \, (\mathbb{P}(|U_{j,x,h}| >$
Then it is

$$\begin{aligned}
&\mathbb{P}\left(\frac{1}{T(n)}\left|\sum_{j=1}^{T(n)} \eta_{j,x,h}\right| \ge c_n, \delta_n^{(1)} \le T(n) \le \delta_n^{(2)}\right) \\
&\quad\le \mathbb{P}\left(\bigcup_{i=1}^{T(n)} A_i\right) \le \sum_{i=1}^{T(n)} \mathbb{P}\left(A_i\right) \le T(n)\mathbb{P}\left(B_1\right) \\
&\quad\le (\delta_n^{(1)})^2 \mathbb{P}\left(|U_{j,x,h}| > R_n\right) \le C(\delta_n^{(1)})^2 h^{(1-q)d} R_n^{-q} \ , \qquad \text{(A.25)}
\end{aligned}$$

where the last inequality follows from Markov inequality and moment bounds on $\mathbb{E}\left(|U_{j,x,h}|^q\right)$ for $q \ge 2$ by Lemma A.1 equation (A.15).

Thus putting (A.22), (A.23),(A.24), and (A.25) together, it is

$$\sum l(n)\mathbb{P}\left(\frac{1}{T(n)}\left|\sum_{i=0}^{n} W_{i,x,h}\right| \geq c_n\right) < \infty\,, \tag{A.26}$$

for appropriate $\delta_n^{(1)}, \delta_n^{(2)}, R_n, h, c_n$. For simplicity set $\delta_n^{(1)} = n^{\beta-\alpha}, \delta_n^{(2)} = n^{\beta+\alpha}, R_n = n^{\beta-\gamma}$. Set $0 < \delta < \alpha < \gamma < \beta$ and γ such that $R_n \gg h^{-d}$. Such a choice is possible due to the mild restriction on h in the assumptions. With $l(n)$ as in defined in (A.19) we can choose q large enough, such that from (A.25) $l(n)(\delta_n^{(1)})^2 h^{(1-q)d} R_n^{-q} < n^{-1}$ in order to fulfill (A.22). With $c_n = O(n^{1/2(\beta-\delta)}h^{d/2}) = O(h^{d+1}l(n)^{1/d})$ we obtain summability in (A.26) of part (A.24) and satisfy the convergence conditions for the terms Q_1 and Q_3.

Therefore with these choices (A.26) holds, resulting in the entire term S_1 being $o(1)$ due to Borel–Cantelli lemma. The final rate combines the rates of the stochastic term S_1 and the bias term $S_2 = O(h^2)$ or $S_2 = O(h)$ on the boundary. □

Remark 5. In the proof above in the case $d = 1$ e.g. the choices $\lambda = \alpha = 1/5, \delta = 1/10$ and $\gamma = 2/5$ are possible. Then $q \geq 10$ allows to satisfy the conditions.

A.3.2 Proof of (A.2)

In order to show (A.2), bias and variance part are treated separately according to the decomposition (A.3). The following two Lemmas can be easily generalized to $x \in \mathbb{R}^d$ and $m : \mathbb{R}^d \to \mathbb{R}$, but here univariate results are sufficient.

Lemma A.3. *Let either Assumptions 1-3 or Assumptions 1,2, and 3* hold. Choose a bandwidth* $h = n^{-\lambda\beta}$ *with* $0 < \lambda < 1 - \frac{\delta+\kappa}{\beta}$ *where* $\kappa > 0$ *arbitrarily small*

$$\sup_{x^j \in \hat{\mathcal{G}}_{jh}} \left|\widehat{m}_j^B(x^j) - m_j(x^j)\right| \quad = \quad O_P\left(h^2\right)\,. \tag{A.27}$$

On the boundary the rate of bias is of order h *instead of* h^2.

Proof. With standard kernel calculations and Taylor expansion it is $\mathbb{E}\left(\widehat{m}_j^B(x^j)|X_1^j, \ldots, X_n^j\right) = m_j(x^j) + R_{n,h}(x^j)$, where $R_{n,h}(x^j) = \frac{r_{n,h}(x^j)}{\widehat{\pi}_j(x^j)}$ and the leading terms in the numerator $r_{n,h}(x^j)$ have the form

$$\begin{aligned} r_{n,h}(x^j) = hm_j^{(1)}(x^j)\frac{1}{T^j(n)}\sum_{i=1}^{n} S_j^{i,1}(x^j) + \\ +h^2 m_j^{(2)}(x^j)\frac{1}{T^j(n)}\sum_{i=1}^{n} S_j^{i,2}(x^j)(1+o(1)) \end{aligned} \tag{A.28}$$

with $S_j^{i,l}(x^j) = K_{h,x^j}(X_i^j)\left(\frac{(X_i^j - x^j)}{h}\right)^l$ for $l \in \{1,2\}$. To obtain the result, the only thing left to show is that for the centered $S_j^{i,l\star}$ with $S_j^{i,l\star}(x^j) = S_j^{i,l}(x^j) - \widehat{\pi}_j(x^j)\mu(S_j^{i,l}(x^j))$ it is

$$\sup_{x^j} \frac{1}{T^j(n)}\sum_{i=1}^{n} S_j^{i,l\star}(x^j) = o_P(1)$$

for $l \in \{1,2\}$. This follows from the previous Lemma A.2 with kernel $u^l K(u)$ instead of $K(u)$. The first term in (A.28) vanishes apart from the boundary and the final result follows with Lemma A.2 in the denominator. □

Lemma A.4. *Let Assumptions 1 - 3 or 1 - 3* hold. Choose a bandwidth $h = n^{-\lambda\beta}$ with $0 < \lambda < 1 - \frac{\delta+\kappa}{\beta}$ where $\kappa > 0$ arbitrarily small, then*

$$\sup_{x^j \in \mathcal{G}_j} \left| \widehat{m}_j^A(x^j) - \mu_{(j\varepsilon)}(K_{x^j,h} \otimes id_\varepsilon) \frac{T_{j\varepsilon}(n)}{\widehat{L}_j(x^j)} \right| \quad = \quad o_P\left(\frac{1}{\sqrt{n^{\beta-\delta}h}} \right) .$$

Proof. The proof follows along the lines of lemma A.17 above, where in the dominating numerator the split chain parts $U_k(K_{x,h} \otimes \mathrm{id}_\varepsilon)$ are for the compound chain (X^j, ϵ). We obtain

$$\sup_{x^j} \frac{1}{T^{j\varepsilon}(n)} \sum_{i=0}^{n} \left(K_{h,x^j}(X_i^j)\epsilon_i - \mu_{(j\varepsilon)}(K_{x^j,h} \otimes \mathrm{id}_\varepsilon) \right) = o_P\left(\frac{1}{\sqrt{n^{\beta-\delta}h}} \right)$$

with the analogous truncation steps as in Lemma A.2 and using the respective moment bound (A.16) in (A.25). As under Assumptions 3 bound (A.16) holds for arbitrary q we can set for the Borel-Cantelli argument to hold

$$q > 1 + \frac{\beta(3+\lambda) - 2\omega - 3\delta + 2}{2\omega}$$

with $\omega = \delta + \kappa - \gamma$ and $\gamma > 0$ in $R_n = n^{\beta-\gamma}$ chosen such that $\omega > 0$ in addition to $0 < \delta < \alpha < \gamma$. As $\delta - \gamma < 0$ it is $\omega < \kappa$. Under mixing residuals in Assumption 3* the same considerations amount to requiring for p in $\mathbb{E}(|\varepsilon|^{p(k+1)}) < \infty$ that

$$p > 1 + \frac{\beta(3+\lambda(1-\frac{1}{k+1})) - 2\omega - 3\delta + 2}{2\omega}$$

with $0 < \omega < \kappa$. Note that under this condition the bandwidth requirement for controlling U_0 in Karlsen, Myklebust, and Tjøstheim (2007) Theorem 5.5 is automatically fulfilled. In both cases the final result follows with Lemma A.2 in the denominator. □

Remarks 6.

1. In general, the stochastic bias term $\mu_{(j\varepsilon)}(K_{x^j,h}(\cdot) \otimes \mathrm{id}_\varepsilon)\frac{T_{j\varepsilon}(n)}{\widehat{L}_j(x^j)}$ is $o_p(1)$ (see (6.23) in Karlsen, Myklebust, and Tjøstheim (2007)). With bandwidth $h < n^{1/5\beta+\delta}$ the term vanishes.

2. Under Assumptions 1,2 and 3, we get for $\widehat{m}_j^{(k)}$ rates with bivariate β_{jk} on $\mathcal{G}_{jk}$ and a stochastic bias $\mu_{(j\varepsilon)}^{(k)}(K_{x^j,h}(\cdot) \otimes \mathrm{id}_\varepsilon)\frac{T_{j\varepsilon}^{(k)}(n)}{\widehat{L}_j^{(k)}(x^j)}$.

3. If ϵ and X^j are independent, or only asymptotically independent, then it is $\pi_{j\varepsilon} = \pi_j \cdot \pi_\varepsilon$. Thus $\mu_{(j\varepsilon)}(K_{x^j,h}(X) \otimes \mathrm{id}_\varepsilon)\frac{T^{j\varepsilon}(n)}{\widehat{L}^j(x^j)} = 0$ under Assumption 3*.

References

Bandi, F., and P. Phillips (2007): "A simple approach to the parametric estimation of potentially nonstationary diffusions," *Journal of Econometrics*, 137(2), 354–395.

Bandi, F. M., and P. C. Phillips (2004): *Handbook of Financial Econometrics*chap. Nonstationary Continuous-Time Processes, pp. 1–63.

Cai, Z., Q. Li, and J. Park (2009): "Functional-coefficient models for nonstationary time series data," *Journal of Econometrics*, 148(2), 101–113.

Carrasco, M., J.-P. Florens, and E. Renault (2003): "Linear Inverse Problems in Structural Econometrics Estimation based on Spectral decomposition and regularization," Discussion paper, IDEI working paper.

Chen, X. (2000): "On the limit laws of the second order for additive functionals of Harris recurrent Markov chains," *Probab. Theory and Relat.Fields*, 116, 89–123.

Christopeit, N., and S. G. N. Hoderlein (2006): "Local Partitioned Regression," *Econometrica*, 74(3), 787–817.

Cline, D., and U. Pu (1999): "Stability of nonlinear AR(1) time series with delay," *Stochastic Processes and their Applications*, 82, 307–333.

Darling, D. A., and M. Kac (1957): "On occupation times of Markoff Processes," *Transactions of the American Mathematical Society*, 84, 444–458.

Engle, R. F., C. W. J. Granger, J. Rice, and A. Weiss (1986): "Semiparametric Estimates of the Relation Between Weather and Electricity Sales," *Journal of the American Statistical Association*, 81(394), 310–320.

Feller, W. (1971): *An Introduction to Probability Theory and Its Applications*, vol. II. Wiley.

Gao, J., M. L. King, Z. Lu, and D. Tjøstheim (2009): "Specification testing in nonstationary time series autoregression," *Annals of Statistics*, 37, 3893–3928.

Grossman, G., and A. B. Krueger (1995): "Economic Growth and the Environment," *The Quarterly Journal of Economics*, 110(2), 353–77.

Guerre, E. (2004): "Design–Adaptive Pointwise Nonparametric Regression Estimation For Recurrent Markov Time Series," .

Hall, P., and C. Heyde (1980): *Martingale Limit Theorems and its Application*. Academic Press, New York.

Harbaugh, W., A. Levinson, and D. M. Wilson (2002): "Reexamining The Empirical Evidence For An Environmental Kuznets Curve," *The Review of Economics and Statistics*, 84(3), 541–551.

Höpfner, R., and E. Löcherbach (2000): "Limit Theorems for null recurrent Markov Processes," .

Karlsen, H., T. Myklebust, and D. Tjøstheim (2007): "Nonparametric Estimation in a Nonlinear Cointegration Type Model," *Annals of Statistics*, 35(1), 1–57.

Karlsen, H., and D. Tjøstheim (1998): "Nonparametric Estimation in Null–Recurrent Time Series," *Working paper HU Berlin.*

——— (2001): "Nonparametric Estimation in Null–Recurrent Time Series," *Annals of Statistics*, 29(2), 372–416.

Lewbel, A., and S. Ng (2005): "Demand Systems with Nonstationary Regressors," *Rev. of Econ. and Stat.*, 87(3), 479–494.

Linton, O., and J. P. Nielsen (1995): "A kernel method of estimating structured nonparametric regression based on marginal integration," *Biometrika*, 82, 93–100.

Mammen, E., and O. Linton (2005): "Estimating semiparametric ARCH(∞) models by kernel smoothing methods,," *Econometrica*, 73(3), 771–836.

Mammen, E., O. Linton, and J. Nielsen (1999): "The Existence and Asymptotic Properties of a Backfitting Projection algorithm under Weak conditions," *Annals of Statistics*, 27(5), 1443–1490.

Mammen, E., and B. U. Park (2005): "Bandwidth selection for smooth backfitting in additive models," *Ann. Statist.*, 33(3), 1260–1294.

Meyn, S., and R. Tweedie (1993): *Markov Chains and Stochastic Stability.* Springer.

Myklebust, T., H. A. Karlsen, and D. Tjøstheim (2010): "Null recurrent unit root processes," *Econometric Theory*, forthcoming.

Nielsen, J. P., and S. Sperlich (2005): "Smooth backfitting in practice," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1), 43–61.

Nummelin, E. (1984): *General Irreducible Markov Chains and Non–negative Operators.* Cambridge University Press.

Phillips, P. C., and J. Y. Park (1998): "Nonstationary Density Estimation and Kernel Autoregression," Cowles Foundation Discussion Papers 1181, Cowles Foundation, Yale University.

Sato, K. (1999): *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press.

Schienle, M. (2008): "Nonparametric Nonstationary Regression," Working Paper Mannheim University.

Taylor, M. P., and L. Sarno (1998): "The behavior of real exchange rates during the post-Bretton Woods period," *Journal of International Economics*, 46(2), 281–312.

Tjøstheim, D., and B. Auestad (1994): "Nonparameteric identification of nonlinear time series: Projections," *J.Amer.Stat.Assoc.*, 89(428), 1398–1409.

Wang, Q., and P. C. B. Phillips (2009a): "Asymptotic Theory for Local Time Density Estimation and Nonparametric Cointegrating Regression," *Econometric Theory*, 25(3), 710–738.

——— (2009b): "Structural Nonparametric Cointegrating Regression," *Econometrica*, 77, 1901–1948.

SFB 649 Discussion Paper Series 2011

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Localising temperature risk" by Wolfgang Karl Härdle, Brenda López Cabrera, Ostap Okhrin and Weining Wang, January 2011.
002 "A Confidence Corridor for Sparse Longitudinal Data Curves" by Shuzhuan Zheng, Lijian Yang and Wolfgang Karl Härdle, January 2011.
003 "Mean Volatility Regressions" by Lu Lin, Feng Li, Lixing Zhu and Wolfgang Karl Härdle, January 2011.
004 "A Confidence Corridor for Expectile Functions" by Esra Akdeniz Duran, Mengmeng Guo and Wolfgang Karl Härdle, January 2011.
005 "Local Quantile Regression" by Wolfgang Karl Härdle, Vladimir Spokoiny and Weining Wang, January 2011.
006 "Sticky Information and Determinacy" by Alexander Meyer-Gohde, January 2011.
007 "Mean-Variance Cointegration and the Expectations Hypothesis" by Till Strohsal and Enzo Weber, February 2011.
008 "Monetary Policy, Trend Inflation and Inflation Persistence" by Fang Yao, February 2011.
009 "Exclusion in the All-Pay Auction: An Experimental Investigation" by Dietmar Fehr and Julia Schmid, February 2011.
010 "Unwillingness to Pay for Privacy: A Field Experiment" by Alastair R. Beresford, Dorothea Kübler and Sören Preibusch, February 2011.
011 "Human Capital Formation on Skill-Specific Labor Markets" by Runli Xie, February 2011.
012 "A strategic mediator who is biased into the same direction as the expert can improve information transmission" by Lydia Mechtenberg and Johannes Münster, March 2011.
013 "Spatial Risk Premium on Weather Derivatives and Hedging Weather Exposure in Electricity" by Wolfgang Karl Härdle and Maria Osipenko, March 2011.
014 "Difference based Ridge and Liu type Estimators in Semiparametric Regression Models" by Esra Akdeniz Duran, Wolfgang Karl Härdle and Maria Osipenko, March 2011.
015 "Short-Term Herding of Institutional Traders: New Evidence from the German Stock Market" by Stephanie Kremer and Dieter Nautz, March 2011.
016 "Oracally Efficient Two-Step Estimation of Generalized Additive Model" by Rong Liu, Lijian Yang and Wolfgang Karl Härdle, March 2011.
017 "The Law of Attraction: Bilateral Search and Horizontal Heterogeneity" by Dirk Hofmann and Salmai Qari, March 2011.
018 "Can crop yield risk be globally diversified?" by Xiaoliang Liu, Wei Xu and Martin Odening, March 2011.
019 "What Drives the Relationship Between Inflation and Price Dispersion? Market Power vs. Price Rigidity" by Sascha Becker, March 2011.
020 "How Computational Statistics Became the Backbone of Modern Data Science" by James E. Gentle, Wolfgang Härdle and Yuichi Mori, May 2011.
021 "Customer Reactions in Out-of-Stock Situations – Do promotion-induced phantom positions alleviate the similarity substitution hypothesis?" by Jana Luisa Diels and Nicole Wiebach, May 2011.

SFB 649, Spandauer Str. 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

SFB 649 Discussion Paper Series 2011

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

022 "Extreme value models in a conditional duration intensity framework" by Rodrigo Herrera and Bernhard Schipp, May 2011.
023 "Forecasting Corporate Distress in the Asian and Pacific Region" by Russ Moro, Wolfgang Härdle, Saeideh Aliakbari and Linda Hoffmann, May 2011.
024 "Identifying the Effect of Temporal Work Flexibility on Parental Time with Children" by Juliane Scheffel, May 2011.
025 "How do Unusual Working Schedules Affect Social Life?" by Juliane Scheffel, May 2011.
026 "Compensation of Unusual Working Schedules" by Juliane Scheffel, May 2011.
027 "Estimation of the characteristics of a Lévy process observed at arbitrary frequency" by Johanna Kappus and Markus Reiß, May 2011.
028 "Asymptotic equivalence and sufficiency for volatility estimation under microstructure noise" by Markus Reiß, May 2011.
029 "Pointwise adaptive estimation for quantile regression" by Markus Reiß, Yves Rozenholc and Charles A. Cuenod, May 2011.
030 "Developing web-based tools for the teaching of statistics: Our Wikis and the German Wikipedia" by Sigbert Klinke, May 2011.
031 "What Explains the German Labor Market Miracle in the Great Recession?" by Michael C. Burda and Jennifer Hunt, June 2011.
032 "The information content of central bank interest rate projections: Evidence from New Zealand" by Gunda-Alexandra Detmers and Dieter Nautz, June 2011.
033 "Asymptotics of Asynchronicity" by Markus Bibinger, June 2011.
034 "An estimator for the quadratic covariation of asynchronously observed Itô processes with noise: Asymptotic distribution theory" by Markus Bibinger, June 2011.
035 "The economics of TARGET2 balances" by Ulrich Bindseil and Philipp Johann König, June 2011.
036 "An Indicator for National Systems of Innovation - Methodology and Application to 17 Industrialized Countries" by Heike Belitz, Marius Clemens, Christian von Hirschhausen, Jens Schmidt-Ehmcke, Axel Werwatz and Petra Zloczysti, June 2011.
037 "Neurobiology of value integration: When value impacts valuation" by Soyoung Q. Park, Thorsten Kahnt, Jörg Rieskamp and Hauke R. Heekeren, June 2011.
038 "The Neural Basis of Following Advice" by Guido Biele, Jörg Rieskamp, Lea K. Krugel and Hauke R. Heekeren, June 2011.
039 "The Persistence of "Bad" Precedents and the Need for Communication: A Coordination Experiment" by Dietmar Fehr, June 2011.
040 "News-driven Business Cycles in SVARs" by Patrick Bunk, July 2011.
041 "The Basel III framework for liquidity standards and monetary policy implementation" by Ulrich Bindseil and Jeroen Lamoot, July 2011.
042 "Pollution permits, Strategic Trading and Dynamic Technology Adoption" by Santiago Moreno-Bromberg and Luca Taschini, July 2011.
043 "CRRA Utility Maximization under Risk Constraints" by Santiago Moreno-Bromberg, Traian A. Pirvu and Anthony Réveillac, July 2011.

SFB 649, Spandauer Str. 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

SFB 649 Discussion Paper Series 2011

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

044 "Predicting Bid-Ask Spreads Using Long Memory Autoregressive Conditional Poisson Models" by Axel Groß-Klußmann and Nikolaus Hautsch, July 2011.
045 "Bayesian Networks and Sex-related Homicides" by Stephan Stahlschmidt, Helmut Tausendteufel and Wolfgang K. Härdle, July 2011.
046 "The Regulation of Interdependent Markets", by Raffaele Fiocco and Carlo Scarpa, July 2011.
047 "Bargaining and Collusion in a Regulatory Model", by Raffaele Fiocco and Mario Gilli, July 2011.
048 "Large Vector Auto Regressions", by Song Song and Peter J. Bickel, August 2011.
049 "Monetary Policy, Determinacy, and the Natural Rate Hypothesis", by Alexander Meyer-Gohde, August 2011.
050 "The impact of context and promotion on consumer responses and preferences in out-of-stock situations", by Nicole Wiebach and Jana L. Diels, August 2011.
051 "A Network Model of Financial System Resilience", by Kartik Anand, Prasanna Gai, Sujit Kapadia, Simon Brennan and Matthew Willison, August 2011.
052 "Rollover risk, network structure and systemic financial crises", by Kartik Anand, Prasanna Gai and Matteo Marsili, August 2011.
053 "When to Cross the Spread: Curve Following with Singular Control" by Felix Naujokat and Ulrich Horst, August 2011.
054 "TVICA - Time Varying Independent Component Analysis and Its Application to Financial Data" by Ray-Bing Chen, Ying Chen and Wolfgang K. Härdle, August 2011.
055 "Pricing Chinese rain: a multi-site multi-period equilibrium pricing model for rainfall derivatives" by Wolfgang K. Härdle and Maria Osipenko, August 2011.
056 "Limit Order Flow, Market Impact and Optimal Order Sizes: Evidence from NASDAQ TotalView-ITCH Data" by Nikolaus Hautsch and Ruihong Huang, August 2011.
057 "Optimal Display of Iceberg Orders" by Gökhan Cebiroğlu and Ulrich Horst, August 2011.
058 "Optimal liquidation in dark pools" by Peter Kratz and Torsten Schöneborn, September 2011.
059 "The Merit of High-Frequency Data in Portfolio Allocation" by Nikolaus Hautsch, Lada M. Kyj and Peter Malec, September 2011.
060 "On the Continuation of the Great Moderation: New evidence from G7 Countries" by Wenjuan Chen, September 2011.
061 "Forward-backward systems for expected utility maximization" by Ulrich Horst, Ying Hu, Peter Imkeller, Anthony Réveillac and Jianing Zhang.
062 "On heterogeneous latent class models with applications to the analysis of rating scores" by Aurélie Bertrand and Christian M. Hafner, October 2011.
063 "Multivariate Volatility Modeling of Electricity Futures" by Luc Bauwens, Christian Hafner and Diane Pierret, October 2011.

SFB 649, Spandauer Str. 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

SFB 649 Discussion Paper Series 2011

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

064 "Semiparametric Estimation with Generated Covariates" by Enno Mammen, Christoph Rothe and Melanie Schienle, October 2011.
065 "Linking corporate reputation and shareholder value using the publication of reputation rankings" by Sven Tischer and Lutz Hildebrandt, October 2011.
066 "Monitoring, Information Technology and the Labor Share" by Dorothee Schneider, October 2011.
067 "Minimal Supersolutions of BSDEs with Lower Semicontinuous Generators" by Gregor Heyne, Michael Kupper and Christoph Mainberger, October 2011.
068 "Bargaining, Openness, and the Labor Share" by Dorothee Schneider, October 2011.
069 "The Labor Share: A Review of Theory and Evidence" by Dorothee Schneider, October 2011.
070 "The Power of Sunspots: An Experimental Analysis" by Dietmar Fehr, Frank Heinemann and Aniol Llorente-Saguer, October 2011.
071 "Econometric analysis of volatile art markets" by Fabian Y. R. P. Bocart and Christian M. Hafner, October 2011.
072 "Financial Network Systemic Risk Contributions" by Nikolaus Hautsch, Julia Schaumburg and Melanie Schienle, October 2011.
073 "Calibration of self-decomposable Lévy models" by Mathias Trabs, November 2011.
074 "Time-Varying Occupational Contents: An Additional Link between Occupational Task Profiles and Individual Wages" by Alexandra Fedorets, November 2011.
075 "Changes in Occupational Demand Structure and their Impact on Individual Wages" by Alexandra Fedorets, November 2011.
076 "Nonparametric Nonstationary Regression with Many Covariates" by Melanie Schienle, November 2011.

SFB 649, Spandauer Str. 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".